ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ÜBER EINEN FABER'SCHEN SATZ

VON PAUL ERDÖS UND GÉZA GRÜNWALD

(Received August 18, 1937)

Die n-te Partialsumme der Fourier-Reihe der geraden Funktion $f(x)$ sei

$$(1) \qquad S_n[f(x)] = a_0 + a_1 \cos\vartheta + a_2 \cos 2\vartheta + \cdots + a_{n-1} \cos (n-1)\vartheta.$$

$$(2) \qquad a_0 = \frac{1}{\pi}\int_0^\pi f(\cos\vartheta)\, d\vartheta, \qquad a_r = \frac{2}{\pi}\int_0^\pi f(\cos\vartheta) \cos r\vartheta\, d\vartheta,$$

$$r = 1, 2, \cdots, n-1.$$

Das zu den Wurzeln

$$(3) \qquad x_k = \cos (2k-1)\frac{\pi}{2n} = \cos\vartheta_k$$

des Tschebyscheffschen Polynoms $T_n(x) = \cos n \arccos x$ gehörige Lagrangesche Interpolationspolynom derselben Funktion hat hingegen die Form[1]

$$(4) \qquad L_n[f(x)] \equiv \sum_{k=1}^{n} f(x_k)\frac{T_n(x)}{T_n'(x_k)(x-x_k)} = c_0 + c_1\cos\vartheta + c_2\cos 2\vartheta + \cdots + c_{n-1}\cos(n-1)\vartheta$$

mit

$$(5) \qquad c_0 = \frac{1}{n}\sum_{k=1}^{n} f(\cos\vartheta_k), \qquad c_r = \frac{2}{n}\sum_{k=1}^{n} f(\cos\vartheta_k)\cos r\vartheta_k, \quad r = 1, 2, \cdots, n-1.$$

Ausser der formellen Übereinstimmung der Formeln (1), (2) und (4), (5) finden sich zwischen den Polynomen (1) und (4) verschiedene wesentliche Analogien, weshalb Faber[2] und neuestens Marczinkiewicz[3] das Verhalten der Polynomfolgen (1) und (4) $n = 1, 2, \cdots$ parallel untersuchten. Bei diesen Untersuchungen gewann Faber folgenden aus dem Gesichtspunkte der Analogie gewissermassen negativen Satz: Es existiert eine in $(0, \pi)$ kontinuierliche, gerade periodische Funktion $f(\cos\vartheta)$ mit der merkwürdigen Eigenschaft, dass

---

[1] L. Fejér, Abschätzung eines Polynoms in einem Intervalle, wenn Schranken für seine Werte und ersten Ableitungswerte in einzelnen Punkten des Intervalls gegeben sind und ihre Anwendung auf die Konvergenzfrage Hermitescher Interpolationsreihen. Math. Zeitschrift 32 (1930) S. 426–457, insb. S. 442.

[2] G. Faber, Über stetige Funktionen (zweite Abhandlung), Math. Annalen 69 (1910) S. 372–443, insb. §9.

[3] I. Marczinkiewicz, Sur l'interpolation I, II, Studia Mathematica VI (1936) S. 1–17, bezw. 67–81.

ihre Fourier Reihe im Intervalle $(0, \pi)$ gleichmässig konvergiert, ihre Lagrange-Interpolationspolynome hingegen, wenn man sie in demselben Intervalle auf die Tschebyscheff-Abscissen (3) bezieht, an unendlich vielen Stellen divergieren.

In dieser Arbeit wollen wir beweisen, dass es eine, im Intervalle $(0, \pi)$ *kontinuierliche, gerade periodische Funktion* $f(\cos \vartheta)$ *gibt so, dass ihre Fourier-Reihe im ganzen Intervalle gleichmässig konvergiert, ihre zu den Tschebyscheffschen Abscissen gehörigen Lagrange-Interpolationspolynome hingegen in einem jeden Punkte des Intervalls* $(0, \pi)$ *unendlich werden.*

Wir bemerken noch, dass unsere kontinuierliche Funktion nirgends differentierbar ist, sie erfüllt sogar keine Lipschitz-Bedingung mit positivem Exponent und dennoch konvergiert ihre Fourier Reihe gleichmässig.[4]

Wir benützen hier diejenige Konstruktion, durch welche einer von uns[5] eine Funktion $f(x)$ angegeben hat, deren Lagrange-Interpolationspolynome überall divergieren. $f(x)$ hatte die Form

$$f(x) = \sum_1^\infty \frac{\varphi_k(x)}{\mu_k}, \tag{6}$$

wo für die Polynome $\varphi_k(x)$ im Intervalle $-1 \leqq x \leqq 1$ die Ungleichungen $|\varphi_k(x)| \leqq 2$ gelten, die μ_k hingegen Zahlen bedeuten, die mit einer durch die Konstruktion angegebenen Schnelligkeit ins Unendliche wachsen und für welche $\sum_1^\infty 1/\mu_k$ konvergent ist. Aus dem Beweis der oben zitierten Arbeit ist es klar, dass die $\varphi_k(x)$ nicht unbedingt Polynome sein müssen, es wird von ihnen nur gefordert:

$$\alpha.\,, |\varphi_k(x)| \leqq 2; \qquad \beta.\,, |L_n[\varphi_k(x)]| \leqq A_k \quad n = 1, 2, \cdots,$$

wo die A_k sowohl von n als auch von x unabhängig sind; $\gamma.\,, \varphi_k(x)$ nimmt in n_k, durch die Konstruktion angegebenen (n_k ganze Zahlen, $k = 1, 2, \cdots$) Stellen, die Werte $+1$ oder 0 an.

Wir benötigen folgendes

LEMMA. *Es seien im Intervalle* $0 \leqq \vartheta \leqq \pi$, $\nu + \mu = n_i$ *Punkte*

$$x_1, x_2, \cdots, x_\nu; \qquad y_1, y_2, \cdots, y_\mu \tag{7}$$

gegeben. Zu einem jeden n_i *kann man eine Lipschitz-Bedingung erfüllende Funktion* $f_i(\cos \vartheta)$ *angeben so, dass*

$$f_i(y_k) = 0, \qquad k = 1, 2, \cdots, \mu, \tag{8}$$

$$f_i(x_k) = 1, \qquad k = 1, 2, \cdots, \nu, \tag{9}$$

$$|f_i(\cos \vartheta)| \leqq 1, \tag{10}$$

$$|S_m[f_i(\cos \vartheta)]| \leqq c_1, \tag{11}$$

---

[4] Es ist leicht zu beweisen, dass, wenn die zu den Tschebyscheffschen Abscissen gehörigen Lagrangeschen Interpolationspolynome der Funktion $f(x)$ im Punkte x_0 divergieren, dann erfüllt die Funktion $f(x)$ im Punkte x_0 mit keinem positiven Exponent eine Lipschitz-Bedingung. Also folgt die Nichtdifferenzierbarkeit einfach aus der Divergenz der Interpolationspolynome.

[5] G. Grünwald, Über Divergenzerscheinungen der Lagrangeschen Interpolationspolynome stetiger Funktionen, Annals of Mathematics 37 (1936) S. 908–918.

wo c_1 (ebenso wie später c_2, $\cdots$) eine absolute, also von m, n_i und x unabhängige Konstante bedeutet.

Aus diesem Lemma folgt unser Satz unmittelbar. Da die zu den Tschebyscheffschen Abscissen gehörigen Lagrange-Interpolationspolynome einer Lipschitz Bedingung erfüllenden Funktion gleichmässig zu der Funktion konvergieren,[6] wird für die Funktion $f_i(\cos\vartheta)$ β. ebenso wie wegen (10), (8) und (9) auch α., und γ., erfüllt, woraus nach unseren früheren Bemerkungen folgt, dass bei entsprechender Wahl der Punkte (7) und der Zahlen μ_k die Lagrange-Interpolationspolynome der stetigen Funktion

$$f(x) = \sum_1^\infty \frac{f_k(x)}{\mu_k} \qquad x = \cos\vartheta$$

überall divergieren.
Anderseits wegen

$$S_m[f(x)] = S_m\left[\sum_1^r \frac{f_k(x)}{\mu_k}\right] + S_m\left[\sum_{r+1}^\infty \frac{f_k(x)}{\mu_k}\right] = S_m[\varphi(x)] + S_m[\psi(x)]$$

erhalten wir wegen (11) für genügend grosses r und gegebenes $\epsilon > 0$

$$|S_m[f(x)] - S_m[\varphi(x)]| \leqq c_1 \sum_{r+1}^\infty 1/\mu_k < \frac{\epsilon}{2},$$

Also, da $\varphi(x)$ eine Lipschitz-Bedingung erfüllt und daher seine Fourier-Reihe gleichmässig konvergiert, gilt für genügend grosses m

$$|S_m[f(x)] - \varphi(x)| < \epsilon,$$

womit unser Satz bewiesen ist. Also haben wir nur unser Lemma zu beweisen.

Wir bezeichnen die kleinste der Entfernungen zweier Nachbarpunkte von (7) durch d und umgeben einen jeden der Punkte x_k mit einem Intervall (α_k, β_k) so, dass die Länge eines solchen Intervalls $\beta_k - \alpha_k = d/n_i$, $k = 1, 2, \cdots \gamma$ ist und x_k genau im Mittelpunkt des Intervalls (α_k, β_k) liegt.

Nun definieren wir $f_i(x)$ folgendermassen:

$f_i(x) = 2(x - \alpha_k)/(\beta_k - \alpha_k)$, wenn $\alpha_k \leqq x \leqq \frac{1}{2}(\alpha_k + \beta_k)$

$f_i(x) = 2(x - \beta_k)/(\alpha_k - \beta_k)$, wenn $\frac{1}{2}(\alpha_k + \beta_k) \leqq x \leqq \beta_k$

$f_i(x) = 0$, wenn $\beta_k \leqq x \leqq \alpha_{k+1}$, $-1 \leqq x \leqq \alpha_1$,

$\beta_\nu \leqq x \leqq +1$.

Es ist klar, dass bei dieser Definition $f_i(x)$ eine Lipschitz-Bedingung und auch (8), (9) und (10) erfüllt, also haben wir nur (11) zu beweisen.

---

[6] Siehe zum B. L. Fejér, Lagrangesche Interpolation und die angehörigen konjugierten Punkte, Math. Annalen 106 (1932) S. 1–55.

Es sei $\alpha_j \leqq x \leqq \alpha_{j+1}$ und es sei vorausgesetzt, dass x näher zu α_j als zu α_{j+1} liegt (x wird jetzt für fest betrachtet). Dann ist

$$
\begin{aligned}
|S_m[f_i(x)]| &= \left|\frac{1}{\pi}\sum_{k=1}^{\nu}\left(\int_{\alpha_k}^{\frac{1}{2}(\alpha_k+\beta_k)} \frac{2(t-\alpha_k)}{\beta_k-\alpha_k}\,\frac{\sin(m+\frac{1}{2})(t-x)}{\sin\frac{1}{2}(t-x)}\,dt\right.\right. \\
&\quad \left.\left. + \int_{\frac{1}{2}(\alpha_k+\beta_k)}^{\beta_k} \frac{2(t-\beta_k)}{\alpha_k-\beta_k}\,(t-\beta_k)\,\frac{\sin(m+\frac{1}{2})(t-x)}{\sin\frac{1}{2}(t-x)}\,dt\right)\right| \\
&\leqq \left|\frac{1}{\pi}\sum_{k=1}^{\nu}\left(\int_{\alpha_k}^{\frac{1}{2}(\alpha_k+\beta_k)} \frac{4(t-\alpha_k)}{\beta_k-\alpha_k}\,\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt\right.\right. \\
&\quad \left.\left. + \int_{\frac{1}{2}(\alpha_k+\beta_k)}^{\beta_k} \frac{4(t-\beta_k)}{\beta_k-\alpha_k}\,\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt\right)\right| \\
&\quad + c_2 = \frac{1}{\pi}\left|\sum^{\nu}(T_k+T_k')\right| + c_1.
\end{aligned} \tag{12}
$$

Für $r \geqq 1$

$$
\begin{aligned}
|T_{j-r}| &= \left|\int_{\alpha_{j-r}}^{\frac{1}{2}(\alpha_{j-r}+\beta_{j-r})} \frac{4(t-\alpha_{j-r})}{\beta_{j-r}-\alpha_{j-r}}\,\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt\right| \\
&< 4\int_{\alpha_{j-r}}^{\frac{1}{2}(\alpha_{j-r}+\beta_{j-r})}\left|\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\right|dt.
\end{aligned}
$$

Da nach unserer Voraussetzung $|t-x| \geqq rd$

$$
|T_{j-r}| < \frac{4}{rd}\int_{\alpha_{j-r}}^{\frac{1}{2}(\alpha_{j-r}+\beta_{j-r})} |\sin(m+\tfrac{1}{2})(t-x)|\,dt < \frac{2(\beta_{j-r}-\alpha_{j-r})}{r\cdot d} < \frac{2}{n_i}\cdot\frac{1}{r}.
$$

Ebenso für $r \geqq 1$, wegen $|t-x| \geqq \frac{1}{2}rd$

$$
|T_{j+r}| < 4/n_i r,
$$

ferner

$$
|T'_{j-r}| < 2/n_i r, \qquad |T'_{j+r}| < 4/n_i r.
$$

Also

$$
\sum_{\substack{1\leqq k\leqq \nu\\ k\neq j}} (|T_k| + |T_k'|) < 2\cdot\frac{4}{n_i}\log\nu < \frac{8\log n_i}{n_i} < c_3. \tag{13}
$$

Da

$$
\begin{aligned}
T_j &= \frac{4}{\beta_j-\alpha_j}\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)} \frac{4(t-\alpha_j)}{\beta_j-\alpha_j}\,\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt \\
&= \int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)} \frac{4}{\beta_j-\alpha_j}\sin(m+\tfrac{1}{2})(t-x)\,dt \\
&\quad + \frac{4(x-\alpha_j)}{\beta_j-\alpha_j}\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)} \frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt,
\end{aligned}
$$

für $|x-\alpha_j| \leqq 4(\beta_j-\alpha_j)$

$$|T_j| \leqq \frac{4}{\beta_j-\alpha_j}\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)} |\sin(m+\tfrac{1}{2})(t-x)|\,dt$$
$$+\frac{4(x-\alpha_j)}{\beta_j-\alpha_j}\left|\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)}\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt\right|$$
$$<2+16\left|\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)}\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\,dt\right|<c_4,$$

und für $|x-\alpha_j|>4(\beta_j-\alpha_j)$ wegen $|t-x|>3(\beta_j-\alpha_j)$

$$|T_j| \leqq 4\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)}\left|\frac{\sin(m+\frac{1}{2})(t-x)}{t-x}\right|dt$$
$$<\frac{4}{3(\beta_j-\alpha_j)}\int_{\alpha_j}^{\frac{1}{2}(\alpha_j+\beta_j)}|\sin(m+\tfrac{1}{2})(t-x)|\,dt<\frac{2}{3};$$

also $|T_j| \leqq \max(\frac{2}{3}, c_4) = c_5$. Ebenso $|T_j'| \leqq c_6$. Also mit Rücksicht auf (12) und (13)

$$|S_m[f_i(x)]| \leqq \frac{1}{\pi}\left|\sum_{k=1}^{\nu}(T_k+T_k')\right|+c_2 \leqq \frac{1}{\pi}(c_3+c_5+c_6)+c_2=c_1,$$

womit alles bewiesen ist.

MANCHESTER, ENGLAND
UND
BUDAPEST, HUNGARY.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ORTHOGONAL POLYNOMIALS IN TWO COMPLEX VARIABLES[1]

BY DUNHAM JACKSON

(Received November 18, 1937)

1. Introduction

Comparison of the theory of orthogonal polynomials in a single complex variable[2] with that of orthogonal polynomials in two real variables[3] readily suggests the primary facts with regard to orthogonal polynomials in two complex variables. Such facts may be enumerated without details of demonstration. The extension from two to four real dimensions, however, naturally increases the variety of the phenomena to be noted, and after the first stages of analogy leads to developments which call for more sustained examination. The purpose of this paper is to give an introduction to the theory in two complex variables, with emphasis on certain features of it which are not brought out in the presentations that have come to the writer's attention.[4]

2. Formal properties

The independent variables will be denoted by (z_1, z_2),

$$z_1 = x_1 + iy_1, \qquad z_2 = x_2 + iy_2.$$

---

[1] Presented to the American Mathematical Society September 8, 1937.

[2] See e.g. G. Szegö, *Beiträge zur Theorie der Toeplitzschen Formen II*, Mathematische Zeitschrift, vol. 9 (1921), pp. 167–190; G. Szegö, *Über orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehören*, ibid., pp. 218–270; T. Carleman, *Über die Approximation analytischer Funktionen durch lineare Aggregate von vorgegebenen Potenzen*, Arkiv för Matematik, Astronomi och Fysik, vol. 17 (1922–23), No. 9; J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 20, New York, 1935, Chapter 6.

[3] See D. Jackson, *Formal properties of orthogonal polynomials in two variables*, Duke Mathematical Journal, vol. 2 (1936), pp. 423–434; D. Jackson, *Orthogonal polynomials on a plane curve*, Duke Mathematical Journal, vol. 3 (1937), pp. 228–236.

[4] See B. Almer, *Sur quelques problèmes de la théorie des fonctions analytiques de deux variables complexes*, Arkiv för Matematik, Astronomi och Fysik, vol. 17 (1922–23), No. 7, pp. 49–66; S. Bergmann, *Über unendliche Hermitesche Formen, die zu einem Bereiche gehören, nebst Anwendungen auf Fragen der Abbildung durch Funktionen von zwei komplexen Veränderlichen*, Mathematische Zeitschrift, vol. 29 (1929), pp. 641–677, especially pp. 658, 660–662; S. Bergmann, *Über die Existenz von Repräsentantenbereichen in der Theorie der Abbildung durch Paare von Funktionen zweier komplexen Veränderlichen*, Mathematische Annalen, vol. 102 (1930), pp. 430–446; S. Bergmann, *Über den Wertevorrat einer Funktion von zwei komplexen Veränderlichen*, Mathematische Zeitschrift, vol. 36 (1933), pp. 171–183; A. Hammerstein, *Über die Approximation von Funktionen zweier komplexer Veränderlicher durch Polynome*, Berliner Sitzungsberichte, Phys.-Math. Klasse, 1933, pp. 259–266.

If the polynomials of the orthogonal system are identified by double subscripts, the first subscript indicating in each case the degree of the polynomial in the two variables together, the relation of orthogonality is of the form

$$\int_D \rho(z_1, z_2) p_{nm}(z_1, z_2) \overline{p_{lk}(z_1, z_2)}\, dE = 0, \qquad |n - l| + |m - k| > 0,$$

the bar denoting the complex conjugate. The domain of integration D may be a more or less arbitrary point set, but for purposes of illustration will be thought of as a bounded one-, two-, three-, or four-dimensional spread of simple form in the space of the four real coördinates (x_1, y_1, x_2, y_2). The weight function $\rho(z_1, z_2)$, not necessarily defined except on D, is real, integrable, non-negative, and positive on a set of positive measure there. The polynomials will be assumed normalized so that the value of the integral is 1 when $n = l$, $m = k$.

If D is of three or four dimensions, any finite number of the monomials

$$1, z_1, z_2, z_1^2, z_1 z_2, z_2^2, z_1^3, \cdots$$

will be linearly independent on D, and the orthogonal system will include $n + 1$ polynomials of the nth degree for each value of n. For a one- or two-dimensional spread this may or may not be the case. A linear relation connecting a finite number of the monomials means *ipso facto* that D is a part of an algebraic locus. (When assumed to be of finite extent it can not be a complete algebraic locus in the complex domain). All polynomials vanishing identically on D (thought of as a closed point set without isolated points) have as a common factor one such polynomial of lowest degree. If this degree is p, the orthogonal system contains just p polynomials of the nth degree for $n \geqq p - 1$.

For any particular value of n let ν_n, or for greater simplicity ν, denote the number of polynomials of the nth degree in the system, $n + 1$ or p as the case may be. The ν polynomials $p_{nk}(z_1, z_2)$ are not significantly determined individually, for any linear combinations of them with coefficients forming a unitary matrix will be likewise normalized, orthogonal to each other, and orthogonal to every polynomial of lower degree. Conversely, however, the definition is determinate to the extent that any two sets of ν polynomials of the nth degree having the specified properties are expressible in terms of each other by unitary transformation. (The real orthogonal transformations associated with the corresponding problem in real variables are naturally replaced here by unitary transformations admitting complex coefficients.) If the domain D and the weight function ρ are invariant under a linear transformation

$$z_1' = \alpha_{11} z_1 + \alpha_{12} z_2, \qquad z_2' = \alpha_{21} z_1 + \alpha_{22} z_2,$$

then, at least under simple supplementary conditions (e.g. if the transformation is assumed to be unitary and D is made up of a finite number of congruent parts which are merely permuted among themselves), the ν polynomials $p_{nk}(z_1', z_2')$ are linearly expressible in terms of the corresponding polynomials in the original variables by means of a unitary matrix of coefficients.

In a number of important special cases the monomials $z_1^h z_2^k$ are orthogonal to each other, and except for normalizing coefficients constitute the system of orthogonal polynomials. This can be immediately verified by explicit integration, for example, with unit weight function, in the case of the "unit cylinder," the four dimensional region D_1 defined by the relations

$$|z_1| \leqq 1, \qquad |z_2| \leqq 1,$$

$$z_1 = r_1(\cos\varphi + i\sin\varphi), \qquad z_2 = r_2(\cos\psi + i\sin\psi),$$

$$0 \leqq \varphi \leqq 2\pi, \quad 0 \leqq \psi \leqq 2\pi, \qquad 0 \leqq r_1 \leqq 1, \quad 0 \leqq r_2 \leqq 1,$$

the differential element being

$$dE = r_1 r_2 dr_1 dr_2 d\varphi d\psi,$$

or for the three-dimensional region

$$D_{2a}: |z_1| \leqq 1, \qquad |z_2| = 1,$$

described by the preceding formulas with $r_2 = 1$, the differential element becoming

$$dE = r_1 dr_1 d\varphi d\psi,$$

or the corresponding region D_{2b} with the subscripts 1 and 2 interchanged, or the region D_2 composed of D_{2a} and D_{2b} together and forming the three-dimensional boundary of D_1 ; or the two-dimensional domain

$$D_3 : |z_1| = 1, \qquad |z_2| = 1, \qquad dE = d\varphi d\psi;$$

or the "unit sphere"

$$D_4 : |z_1|^2 + |z_2|^2 \leqq 1,$$

$$z_1 = r\cos\theta(\cos\varphi + i\sin\varphi), \qquad z_2 = r\sin\theta(\cos\psi + i\sin\psi),$$

$$0 \leqq \theta \leqq \pi/2, \qquad 0 \leqq \varphi \leqq 2\pi, \qquad 0 \leqq \psi \leqq 2\pi, \qquad 0 \leqq r \leqq 1,$$

with

$$dE = r^3 \cos\theta \sin\theta dr d\theta d\varphi d\psi;$$

or D_5 , the three-dimensional surface of the sphere, obtained by setting $r = 1$ in the formulas for D_4 , while

$$dE = \cos\theta \sin\theta d\theta d\varphi d\psi.$$

More generally, D_4 can be replaced by[5]

$$D_6 : z_1 = rg_1(\theta)(\cos\varphi + i\sin\varphi), \; z_2 = rg_2(\theta)(\cos\psi + i\sin\psi),$$

$$dE = r^3 g_1(\theta)\, g_2(\theta) \,|\, g_1(\theta)\, g_2'(\theta) - g_2(\theta)\, g_1'(\theta) \,|\, dr d\theta d\varphi d\psi,$$

[5] Cf. S. Bergmann, Mathematische Zeitschrift, vol. 29, loc. cit., pp. 660–661.

where $g_1(\theta)$ and $g_2(\theta)$ are arbitrary functions, for simplicity non-negative and continuous, defined over some interval of values of θ; and D_5 by D_7, consisting of those points of D_6 for which $r = 1$, with

$$dE = g_1(\theta)\, g_2(\theta)\, \{[g_1'(\theta)]^2 + [g_2'(\theta)]^2\}^{\frac{1}{2}} d\theta d\varphi d\psi.$$

In comparing the last two differential elements it is to be noted by way of explanation that in this case the parametric curves for variation of r and θ separately are not orthogonal to each other in general, the angle between them being a function of θ, and the differential of arc in the direction of varying r also depends on θ.

Instead of unity, the weight function may be an arbitrary function of r_1 and r_2 in the case of D_1, or of r_1 on D_{2a}, or of r_1 and r_2 respectively on the corresponding parts of D_2, or of r and θ on D_4 or D_6, or of θ on D_5 or D_7, subject each time to the condition of being integrable, non-negative, and positive on a set of positive measure. The orthogonal polynomials will still be the monomials $z_1^h z_2^k$.

It may be noted in passing that in the case of the one-dimensional locus

$$|z_1| = 1, \qquad z_1 = \bar{z}_2,$$

$$z_1 = \cos\theta + i \sin\theta, \qquad z_2 = \cos\theta - i \sin\theta,$$

which is algebraic of the second degree, being a part of the locus $z_1 z_2 = 1$, the orthogonal polynomials with constant weight function reduce on the curve to the sines and cosines of integral multiples of θ, and the corresponding type of series expansion of an arbitrary function is the Fourier series. It is of course possible to bring the Fourier series into the present setting in more trivial fashion by associating it with a circle in the real plane $y_1 = y_2 = 0$.

3. Convergence

As in other problems of similar character, a relation can be established, at least in the case of the simpler regions, between convergence of the development of a given function in series of the orthogonal polynomials and the degree of approximation attainable by other polynomials. It is perhaps unnecessary to present here a proof of the least-square property[6] of the partial sums of the series, on which the treatment of convergence is based.[7]

Consider first the region D_5, the discussion of which seems to involve a point of detail not encountered in connection with similar problems in one variable. Let $f(z_1, z_2)$ be an arbitrary continuous function on D_5, and $P_n(z_1, z_2)$ the partial sum of its development in series of orthogonal polynomials through terms of the nth degree in the two variables together, the weight function being

[6] Cf. e. g. J. L. Walsh, op. cit., pp. 111–112.

[7] The results can be generalized, without reference to orthogonal polynomials, to apply when the square of the modulus of the error is replaced by the mth power, $m > 0$, in the integral to be minimized.

supposed henceforth to have a positive lower bound throughout the domain in question. Let ϵ_n be a positive number for which there exists a polynomial[8] $p_n(z_1, z_2)$ of the nth degree such that

$$|f(z_1, z_2) - p_n(z_1, z_2)| \leqq \epsilon_n$$

on D_5. Let

$$f(z_1, z_2) - p_n(z_1, z_2) = r_n(z_1, z_2),$$
$$P_n(z_1, z_2) - p_n(z_1, z_2) = \pi_n(z_1, z_2),$$

and let μ_n be the maximum of $|\pi_n(z_1, z_2)|$ on D_5, taken on at a point (z_{10}, z_{20}) corresponding to the values $\theta_0, \varphi_0, \psi_0$ of the parameters.

There would be an apparent complication in the subsequent reasoning if θ_0 had the value 0 or $\pi/2$, making the factor $\cos\theta \sin\theta$ in the differential element vanish, or if as n becomes infinite a succession of values of θ_0 were to come into consideration approaching 0 or $\pi/2$. Since however the essential properties of $P_n(z_1, z_2)$ as the polynomial of the nth degree approximating $f(z_1, z_2)$ in the sense of least squares are not affected by unitary transformation of the variables z_1, z_2, corresponding to an orthogonal transformation in the four-dimensional real space of their components, it is permissible to apply a transformation of this sort for each value of n separately so that θ_0 shall not be near one of the exceptional values, and in fact shall be precisely $\pi/4$. The existence of such a transformation may be recognized as follows.

A unitary transformation of z_1, z_2 is defined by the formulas

$$z_1' = \alpha_{11} z_1 + \alpha_{12} z_2, \qquad z_2' = \alpha_{21} z_1 + \alpha_{22} z_2,$$

with

$$\alpha_{11} = \cos\alpha(\cos\beta + i\sin\beta), \qquad \alpha_{12} = \sin\alpha[\cos(\beta+\delta) + i\sin(\beta+\delta)],$$
$$\alpha_{21} = -\sin\alpha(\cos\gamma + i\sin\gamma), \qquad \alpha_{22} = \cos\alpha[\cos(\gamma+\delta) + i\sin(\gamma+\delta)],$$

the angles $\alpha, \beta, \gamma, \delta$ being arbitrary, except that α may without loss of generality be taken as belonging to the first quadrant. If

$$z_1 = \cos\theta\ (\cos\varphi + i\sin\varphi), \qquad z_2 = \sin\theta\ (\cos\psi + i\sin\psi),$$
$$z_1' = \cos\theta'\ (\cos\varphi' + i\sin\varphi'), \qquad z_2' = \sin\theta'\ (\cos\psi' + i\sin\psi'),$$

it is found that

$$\cos^2\theta' = \cos^2\alpha\cos^2\theta + \sin^2\alpha\sin^2\theta + 2\cos\alpha\sin\alpha\cos\theta\sin\theta\cos(\varphi - \delta - \psi).$$

Let $\theta_0, \varphi_0, \psi_0$ be as specified above, let $\theta_0', \varphi_0', \psi_0'$ be the corresponding values of the accented parameters, and in particular let $\delta = \varphi_0 - \psi_0 - \pi/2$. Then

$$\cos^2\theta_0' = \cos^2\alpha\cos^2\theta_0 + \sin^2\alpha\sin^2\theta_0.$$

[8] This notation p_n with a single subscript is independent of the earlier notation with two subscripts for the individual polynomials of the orthogonal system.

If α is allowed to vary from 0 to $\pi/2$, $\cos^2 \theta_0'$ changes continuously from $\cos^2 \theta_0$ to $\sin^2 \theta_0$, and passes through the value $\frac{1}{2}$ (if not constantly equal to $\frac{1}{2}$). It is possible then to choose the unitary transformation so that $\theta_0' = \pi/4$. For economy of notation it may be supposed without loss of generality that $\theta_0 = \pi/4$ in the first place.

The function $\pi_n(z_1, z_2)$, being a polynomial in z_1 and z_2, is on D_5 a trigonometric sum in each of the variables θ, φ, ψ. It has μ_n as an upper bound for its absolute value, not only for all values of φ and ψ, but also for all values of θ, without restriction to the interval $(0, \pi/2)$, since z_1 or z_2 remains unchanged if $\cos\theta$ or $\sin\theta$ is reversed in sign and φ or ψ at the same time increased by π. Hence the modulus of each of the partial derivatives $\partial\pi_n/\partial\theta$, $\partial\pi_n/\partial\varphi$, $\partial\pi_n/\partial\psi$ has $n\mu_n$ as an upper bound.[9] From this, together with the fact that $\cos\theta \sin\theta$ has a positive lower bound in the neighborhood of the point $(\theta_0, \varphi_0, \psi_0)$, it follows that $|\pi_n(z_1, z_2)|$ remains greater than $\mu_n/2$ throughout a region whose extent, measured by the integral of $\cos\theta \sin\theta d\theta d\varphi d\psi$, is not less than a quantity of the order of magnitude of $1/n^3$. A type of argument which has been applied to numerous other problems of the same sort[10] shows that $|f(z_1, z_2) - P_n(z_1, z_2)|$ will have a constant multiple of $n^{3/2}\epsilon_n$ as an upper bound on D_5. If polynomials $p_n(z_1, z_2)$ exist so that $\lim_{n\to\infty} n^{3/2}\epsilon_n = 0$, $P_n(z_1, z_2)$ converges uniformly on D_5 toward $f(z_1, z_2)$. In particular the condition is surely satisfied if $f(z_1, z_2)$ is analytic throughout D_4, inclusive of the boundary, the partial sums of the double power series expansion serving as the polynomials $p_n(z_1, z_2)$. By the maximum-minimum theorem convergence takes place throughout the interior of the sphere also, i.e. throughout the four-dimensional region D_4.

In connection with the convergence problem that arises when orthogonality and least-square approximation are defined with respect to integration throughout the volume of the four-dimensional sphere, the following form of Bernstein's theorem is sufficient for the purpose in hand:

If $P_n(z_1, z_2)$ is a polynomial of the nth degree in the two variables together such that

$$|P_n(z_1, z_2)| \leqq L$$

for $|z_1|^2 + |z_2|^2 \leqq 1$, then

$$|\partial P_n/\partial z_1| \leqq 2nL, \qquad |\partial P_n/\partial z_2| \leqq 2nL$$

throughout the same region.

The proposition is naturally changed only in form, not in content, if the four-dimensional sphere in the statement is replaced by its three-dimensional boundary.

---

[9] For the admissibility of complex coefficients in the application of Bernstein's theorem cf. S. Bernstein, *Leçons sur les Propriétés Extrêmales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle*, Paris, 1926, pp. 44–45.

[10] See e.g. D. Jackson, *Certain problems of closest approximation*, Bulletin of the American Mathematical Society, vol. 39 (1933), pp. 889–906.

In the first place, the hypothesis implies that $| P_n(z_1 , z_2) | \leqq L$ for $| z_1 | \leqq 2^{-\frac{1}{4}}$, $| z_2 | \leqq 2^{-\frac{1}{4}}$. Hence, by Bernstein's theorem for polynomials in a single complex variable,[11]

$$| \partial P_n/\partial z_1 | \leqq 2^{\frac{1}{4}} nL, \qquad | \partial P_n/\partial z_2 | \leqq 2^{\frac{1}{4}} nL$$

throughout this region, and in particular at the points where $| z_1 | = 2^{-\frac{1}{4}}$, $| z_2 | = 2^{-\frac{1}{4}}$, i.e. at all points of the boundary D_5 for which $\theta = \pi/4$. But it has been seen that any point of D_5 satisfies this condition with respect to a suitably chosen coördinate system, i.e. there exists for any such point a pair of variables z_1', z_2', related to z_1 , z_2 by a unitary transformation, such that

$$| \partial P_n/\partial z_1' | \leqq 2^{\frac{1}{4}} nL, \qquad | \partial P_n/\partial z_2' | \leqq 2^{\frac{1}{4}} nL,$$

from which follows the truth of the proposition as stated.

With the aid of this theorem, applied to the least-square approximation of a function analytic throughout the interior of D_4 and continuous on the boundary, it is possible to arrive at a conclusion with regard to convergence analogous to that obtained above for approximation on D_5 , the factor $n^{3/2}$ being replaced now by n^2, in consequence of the increase in the dimensionality of the domain of integration from three to four. Corresponding results for D_1 , D_2 and D_3 are obtained by more immediate transition from the case of a single complex variable, since the question of rotating axes to avoid singularities in the parametric representation does not arise. The circles can also be replaced by curves and areas of more general shape as ranges for the two variables separately, with the use of appropriate forms of Bernstein's and Markoff's theorems for polynomials in one variable.[12]

THE UNIVERSITY OF MINNESOTA,
MINNEAPOLIS

---

[11] See S. Bernstein, loc. cit.

[12] See D. Jackson, *On certain problems of approximation in the complex domain*, Bulletin of the American Mathematical Society, vol. 36 (1930), pp. 851–857; *On the application of Markoff's theorem to problems of approximation in the complex domain*, the same Bulletin, vol. 37 (1931), pp. 883–890; W. E. Sewell, *A note on the relation between integral and Tchebycheff approximation by polynomials in the complex domain*, the same Bulletin, vol. 43 (1937), pp. 425–431; and other references in the last-named paper.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

REPRESENTATIONS FOR ENTIRE FUNCTIONS OF EXPONENTIAL TYPE

BY R. P. BOAS, JR.[1]

(Received December 7, 1937)

1. Introduction

A theorem of R. E. A. C. Paley and N. Wiener,[2] which has already become classical,[3] states that the entire function $f(z)$ has the representation

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt, \tag{1.1}$$

with $a(t)$ of class $L^2(-R, R)$,[4] if and only if $f(z)$ satisfies

$$|f(z)| < Ae^{R|z|} \tag{1.2}$$

for some constant A, and belongs to L^2 along the real axis.[5] We shall obtain analogues of this theorem for other classes of functions; in particular, for classes L^p $(p \neq 2)$. As one would expect from the theory of Fourier series, the theorem of Paley and Wiener corresponds to a pair of complementary theorems for $1 < p < 2$, each of which fails for $p > 2$. We prove, in fact, that if $f(z)$ satisfies (1.2) and belongs to L^p $(1 < p \leqq 2)$ on the real axis, then $f(z)$ has the form (1.1) with $a(t)$ belonging to $L^{p/(p-1)}(-R, R)$; and that if $f(z)$ has the form (1.1) with $a(t)$ of $L^p(-R, R)$ $(1 < p \leqq 2)$, then $f(z)$ satisfies (1.2) and belongs to $L^{p/(p-1)}$ along the real axis; we construct examples to show that these theorems are both false for $p > 2$.

We also show that $f(z)$ satisfies (1.2) and belongs to L^1 on the real axis if and only if $f(z)$ has the form (1.1) where $a(t)$ has an absolutely convergent Fourier series on $(-R, R)$ and $a(-R+) = a(R-) = 0$; that $f(x)$ belongs to L^1 if it has the form in question is a result of N. Wiener.[6] Furthermore, we show that if $f(z)$ satisfies (1.2) and belongs to L^p $(p > 2)$ on the real axis, then $f(z)$ has the form

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t), \tag{1.3}$$

---

[1] National Research Fellow.

[2] [6], p. 13.

[3] J. D. Tamarkin, review of [7] in Zentralblatt für Mathematik, 16 (1937), p. 360.

[4] $L^p(a, b)$ denotes the class of functions $f(x)$, measurable on (a, b), with $|f(x)|^p$ integrable on (a, b).

[5] The theorem as stated by Paley and Wiener has $|f(z)| = o(e^{R|z|})$, which is a little stronger or a little weaker than (1.2), depending on the half of the theorem on which one's attention is fixed. The difference between the two conditions is unimportant.

[6] [12], p. 80.

where $\alpha(t)$ is continuous, and satisfies a Lipschitz condition of every order less than $1/p$.

After this paper had been completed, I found that the theorem that $f(z)$, satisfying (1.2) and belonging to L^p $(1 \leqq p < 2)$ on the real axis, has the form (1.1) with $a(t)$ as specified above had already been established by M. Plancherel and G. Pólya;[7] the results of these authors concern functions of n variables and are more general than mine in other respects.[7a] However, my method of establishing representations of the form (1.1) or (1.3) is different both from that of Paley and Wiener and that of Plancherel and Pólya, and is independent of the theory of Fourier transforms (which is perhaps a dubious advantage). My point of view is somewhat different from that of Plancherel and Pólya, and some of my auxiliary results are interesting in themselves. In particular, I apply one of them to generalize results of S. Bernstein, C. L. Siegel, Paley and Wiener, and Plancherel and Pólya concerning entire functions of zero type (those satisfying (1.2) for every $R > 0$). If $f(z)$ is of zero type, $f(z)$ cannot be bounded on a line[8] unless it is constant (Bernstein),[9] or belong to L^1 on a line (Siegel),[10] or to L^2 on a line (Paley and Wiener), or to L^p $(1 < p < 2)$ on a line (Plancherel and Pólya),[11] without vanishing identically;[11a] in §5 I show that if $\varphi(t)$ is non-negative, strictly increasing, and convex for $t \geqq 0$, $\varphi(|f(z)|)$ cannot be integrable on a line unless $f(z) \equiv 0$; and that for any positive integer n, $f(z)/(1 + |z|^n)$ belongs to L^p $(p \geqq 1)$ on a line only if $f(z)$ is a polynomial of degree at most $n - 1$ ($n - 2$ if $p = 1$).

2. Preliminary theorems

Throughout this paper we write $z = x + iy = re^{i\theta}$.

DEFINITION.[12] *A function $f(z)$ is an entire function of exponential type R $(R > 0)$ (abbreviation: "$f(z)$ is of type R") if (and only if) $f(z)$ is entire and there exists a positive constant A such that for all z*

$$|f(z)| < Ae^{R|z|}. \tag{2.1}$$

The function $f(z)$ is of zero type if it is of type R for every $R > 0$.

For our purposes, this definition is more convenient than the more precise definition according to which $f(z)$ is of type R if and only if (2.1) is true for R,

---

[7] [7]; see in particular pp. 243, 229 ff.

[7a] Reference should also be made to the paper of Hua and Shü [15]; these authors show that $f(z)$ has the form (1.1) with $a(t)$ of class $L^p(-R, R)$ $(p > 2)$ if and only if $f(z)$ belongs on the real axis to the class H^pL^p of A. C. Offord (See [16] for the definition of H^pL^p).

[8] "A line" is a straight line unlimited in both directions.

[9] [1], p. 77.

[10] [9], p. 323.

[11] [7], p. 245.

[11a] Hua and Shü [15] establish the corresponding theorem for $f(z)$ of class H^pL^p $(p > 2)$ on a line. The introduction of H^p is unnecessary, as I show.

[12] [10], pp. 276, 287.

but fails for every $R' < R$. Under our definition, a function of type R is also of type S for every $S > R$.

LEMMA 2.1. *If $\alpha(t)$ is a bounded (complex-valued) function, such that the integral*

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t) \tag{2.2}$$

exists, then $f(z)$ is of type R.

Integrating (2.2) by parts, we have

$$f(z) = \alpha(R)e^{izR} - \alpha(-R)e^{-izR} - iz\int_{-R}^{R} \alpha(t)e^{izt}\, dt,$$

the sum of three entire functions. Now,

$$\left| z\int_{-R}^{R} \alpha(t)e^{izt}\, dt \right| \leqq A\,|z| \int_{-R}^{R} e^{-yt}\, dt, \qquad A = \sup_{-R\leqq t\leqq R} |\alpha(t)|;$$

for $|y| \geqq \frac{1}{2}|z|$,

$$\left| z\int_{-R}^{R} \alpha(t)e^{izt}\, dt \right| \leqq A\,|z|\,|y|^{-1}(e^{R|y|} - e^{-R|y|})$$

$$\leqq 4Ae^{R|z|};$$

on the other hand, for $|y| < \frac{1}{2}|z|$,

$$\left| z\int_{-R}^{R} \alpha(t)e^{izt}\, dt \right| \leqq 2AR\,|z|\, e^{|y|R} < 2AR\,|z|\, e^{R|z|/2} < Be^{R|z|}$$

for some constant B. We have thus exhibited $f(z)$ as the sum of three functions of type R; such a function is clearly of type R.[13]

LEMMA 2.2 (Valiron).[14] *If $f(z)$ is entire; if for some integer $k \geqq 0$*

$$|f(re^{i\theta})| < \epsilon(r)(1 + r^k)e^{Rr} \qquad (0 \leqq r < \infty), \tag{2.3}$$

where $\lim_{r\to\infty} \epsilon(r) = 0$; and if $f(n\pi/R) = 0$ $(n = 0, \pm 1, \pm 2, \cdots)$, then

$$f(z) = P_{k-1}(z) \sin Rz,$$

where $P_{k-1}(z)$ is a polynomial of degree at most $k - 1$.[15]

3. Necessary and sufficient conditions

THEOREM 3.1. *A necessary and sufficient condition that $f(z)$ have the form*

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt, \tag{3.1}$$

---

[13] A proof for $\alpha(t)$ of bounded variation can be given much more simply.

[14] [11], p. 204.

[15] "$f(z)$ is a polynomial of degree at most m" for $m < 0$ means "$f(z) \equiv 0$".

with $a(t)$ of class $L^p(-R, R)$ $(p \geqq 1)$ is that $f(z)$ be of type R; that

$$\lim_{|n|\to\infty} f(x_n) = 0, \tag{3.2}$$

where the x_n $(n = \pm 1, \pm 2, \cdots)$ are points such that

$$\liminf_{|n|\to\infty} |x_n - n\pi/R| > 0; \tag{3.3}$$

and that for $n = 0, \pm 1, \pm 2, \cdots,$

$$f(n\pi/R) = \int_{-R}^{R} e^{in\pi t/R} a(t)\, dt, \qquad a(t) \in L^p(-R, R). \tag{3.4}$$

Condition (3.4) states that the numbers $f(-n\pi/R)/(2R)$ are the complex Fourier coefficients on $(-R, R)$ of a function of class L^p.

A simpler set of necessary and sufficient conditions would be $f(z)$ of type R, (3.4), and $\lim_{|x|\to\infty} f(x) = 0$; the conditions as we have stated them are more convenient for the applications which we shall make.

As a by-product, Theorem 3.1 gives us an inversion formula for integrals of the form (3.1): the Fourier series

$$\sum_{n=-\infty}^{\infty} \frac{1}{2R} f\left(\frac{-n\pi}{R}\right) e^{in\pi t/R} \tag{3.5}$$

will represent $a(t)$ in any of the usual senses; in particular, (3.5) will be summable $(C, 1)$ to $a(t)$ almost everywhere, and will converge to $a(t)$ if $f(z)$ belongs to L^1 on the real axis (as will follow from Lemma 6.3).

If $f(z)$ has the form (3.1), it is of type R by Lemma 2.1; (3.4) is trivially satisfied; (3.2) is a consequence of the Riemann-Lebesgue theorem.[16]

To show that the conditions are sufficient, we form the function

$$g(z) = \int_{-R}^{R} a(t) e^{izt}\, dt,$$

replace $a(t)$ by its Fourier series (3.5), and integrate term by term. We find

$$g(z) = \sum_{n=-\infty}^{\infty} \frac{1}{2R} f\left(\frac{-n\pi}{R}\right) \int_{-R}^{R} e^{it(z+n\pi/R)}\, dt,$$

$$g(n\pi/R) = f(n\pi/R) \qquad (n = 0, \pm 1, \pm 2, \cdots). \tag{3.6}$$

Because the conditions of the theorem are necessary, Lemma 2.2, with $k = 1$, may be applied to the function $f(z) - g(z)$; hence

$$f(z) - g(z) = A \sin Rz.$$

But $f(z) - g(z)$ satisfies (3.2) for x_n satisfying (3.3); we can choose a sequence $\{x_{n_k}\}$ with $|x_{n_k} - n_k\pi/R| > \delta > 0$, and hence with

$$|\sin Rn_k| > \sin R\delta > 0;$$

consequently $A = 0$. Hence $f(z) = g(z)$, and Theorem 3.1 is established.

---

[16] [14], p. 18.

There are analogous results for the representation of functions in the form

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t). \tag{3.7}$$

If $f(z)$ has the form (3.7), we have

$$f(z) = \alpha(R)e^{izR} - \alpha(-R)e^{-izR} - iz\int_{-R}^{R} e^{izt}\alpha(t)\, dt; \tag{3.8}$$

conversely, if $f(z)$ has the form

$$f(z) = A_1 e^{izR} + A_2 e^{-izR} - iz\int_{-R}^{R} e^{izt}\beta(t)\, dt,$$

$f(z)$ can be written in the form (3.7), where $\alpha(t)$ is obtained from $\beta(t)$ by suitable re-definition at $t = \pm R$. These relations lead us to

THEOREM 3.2. *A necessary and sufficient condition that $f(z)$ have the form (3.7) with $\alpha(t)$ continuous on $-R < t < R$ or of bounded variation on $-R \leqq t \leqq R$ is that $f(z)$ be of type R and that*

$$\frac{1}{2in\pi} f\left(\frac{-n\pi}{R}\right) = \frac{1}{2R}\int_{-R}^{R} e^{-in\pi t/R}\gamma(t)\, dt \qquad (n = \pm 1, \pm 2, \cdots), \tag{3.9}$$

where $\gamma(t)$ is, respectively, continuous on $-R \leqq t \leqq R$ or of bounded variation on $-R \leqq t \leqq R$.

Condition (3.9) states that[17]

$$\sum_{n=-\infty}^{\infty}{}' \frac{1}{2in\pi} f\left(\frac{-n\pi}{R}\right) e^{in\pi t/R}. \tag{3.10}$$

is the Fourier series of a function $\gamma(t)$ of the class in question.

To establish the necessity of the conditions, we write $f(z)$ in the form (3.8), and note that $\alpha(R) - \alpha(-R) = f(0)$, so that

$$\frac{Rf(-n\pi/R)}{in\pi} = \frac{(-1)^n Rf(0)}{in\pi} + \int_{-R}^{R} e^{-in\pi t/R}\alpha(t)\, dt;$$

since

$$\int_{-R}^{R} te^{-in\pi t/R}\, dt = 2(-1)^n R^2/(in\pi),$$

we obtain (3.9) with

$$\gamma(t) = \alpha(t) + \frac{tf(0)}{2R} - \frac{1}{2R}\int_{-R}^{R}\alpha(t)\, dt.$$

To establish the sufficiency of the conditions, we form the function

$$g(z) = \int_{-R}^{R} e^{izt}\, d\gamma(t) = e^{izR}\gamma(R) - e^{-izR}\gamma(-R) - iz\int_{-R}^{R} e^{izt}\gamma(t)\, dt,$$

[17] The accent on the summation sign indicates the omission of the term for which $n = 0$.

replace $\gamma(t)$ by its Fourier series (3.10), and integrate term by term. We find

$$g(z) = e^{izR}\gamma(R) - e^{-izR}\gamma(-R) - iz \sum_{n=-\infty}^{\infty}{}' \frac{f(-n\pi/R)}{2in\pi} \int_{-R}^{R} e^{it(z-n\pi/R)}\,dt,$$

and hence

$$g(n\pi/R) = f(n\pi/R) + (-1)^n G \quad (n = \pm 1, \pm 2, \cdots),$$
$$g(0) = G,$$

where $G = \gamma(R) - \gamma(-R)$. We define a function $h(z)$ by

$$h(z) = g(z) - G\cos Rz + f(0)(Rz)^{-1}\sin Rz.$$

Then $h(z)$ is of type R, and $h(n\pi/R) = f(n\pi/R)$ $(n = 0, \pm 1, \pm 2, \cdots)$. By Lemma 2.2,

$$f(z) - h(z) = A\sin Rz,$$

or

$$(3.11)\quad f(z) = \int_{-R}^{R} e^{izt}\,d\gamma(t) - G\cos Rz + A\sin Rz + (Rz)^{-1}f(0)\sin Rz.$$

If we then define

$$\lambda(t) = \begin{cases} \frac{1}{2}(iA - G), & t = -R; \\ 0, & -R < t < R; \\ \frac{1}{2}(iA + G), & t = R; \end{cases}$$
$$\alpha(t) = \gamma(t) + \lambda(t) + tf(0)/(2R);$$

we may write (3.11) in the form

$$f(z) = \int_{-R}^{R} e^{izt}\,d\alpha(t),$$

where $\alpha(t)$ has bounded variation on $-R \leqq t \leqq R$ if $\gamma(t)$ has, and $\alpha(t)$ is continuous on $-R < t < R$ if $\gamma(t)$ is. This completes the proof.

4. Derivatives of a function of type R

Theorems 3.1 and 3.2, while furnishing necessary and sufficient conditions, are little more than disguised tautologies, and are not easy to apply. We shall obtain conditions which, while they are not in general both necessary and sufficient, are much simpler in appearance. For this purpose we need several lemmas on the derivatives of functions of positive type; incidentally, we shall obtain some information concerning functions of zero type.

LEMMA 4.1. *If $f(z)$ is entire, and if for all z*

$$(4.1)\qquad |f(z)| < |L(z)|\,e^{R|y|},$$

where R is a positive constant and $L(z)$ is a linear function, $Az + B$, then for any $S > R$

$$\frac{d}{dz}\frac{f(z)}{\sin Sz} = -S \lim_{N\to\infty} \sum_{n=-N}^{N} \frac{(-1)^n f(n\pi/S)}{(Sz - n\pi)^2}; \tag{4.2}$$

and

$$\frac{d^2}{dz^2}\frac{f(z)}{\sin Rz} = 2R^2 \sum_{n=-\infty}^{\infty} \frac{(-1)^n f(n\pi/R)}{(Rz - n\pi)^3}. \tag{4.3}$$

A proof may be given by a slight modification of the argument by which G. Pólya and G. Szegö establish (4.2) with $S = R$ when $L(z)$ is a constant.[18]

LEMMA 4.2. (Phragmén-Lindelöf.)[19] *If $f(z)$ is analytic and satisfies*

$$|f(z)| < Ae^{B|z|} \tag{4.4}$$

in the region $\theta_0 < \theta < \theta_0 + \alpha$, where A and B are positive constants, and $0 < \alpha < \pi$; if $f(z)$ is continuous for $\theta_0 \leqq \theta \leqq \theta_0 + \alpha$; and if $|f(z)| \leqq M$ for $\theta = \theta_0$ and $\theta = \theta_0 + \alpha$ (M a positive constant), then $|f(z)| \leqq M$ for $\theta_0 \leqq \theta \leqq \theta_0 + \alpha$.

LEMMA 4.3. *If $f(z)$ is of type R and if*

$$|f(x)| < C(1 + |x|) \qquad (-\infty < x < \infty), \tag{4.5}$$

then $f(z)$ satisfies (4.1), *and* (4.2), (4.3) *are true.*

We consider the function

$$g(z) = [f(z) - f(0)]e^{iRz}z^{-1},$$

which satisfies the hypotheses of Lemma 4.2 in the regions $0 \leqq \theta \leqq \pi/2$ and $\pi/2 \leqq \theta \leqq \pi$. It follows that $g(z)$ is bounded for $y \geqq 0$, or that

$$|f(z)| < M(1 + |z|)e^{Ry} \qquad (y \geqq 0)$$

for some constant M. If we then consider, for $y \leqq 0$,

$$[f(z) - f(0)]e^{-iRz}z^{-1},$$

we obtain a similar inequality for negative y, and hence obtain (4.1).

LEMMA 4.4. (Hardy-Littlewood.)[20] *If $f(x)$ is of class C^2 on $(-\infty, \infty)$ and if, as $x \to \infty$, $f(x) = O(x)$ and $f''(x) = O(x)$, then $f'(x) = O(x)$.*

We can now generalize a theorem of S. Bernstein. This states

LEMMA 4.5.[21] *If $f(z)$ is of type R and $|f(z)| \leqq M$ on the real axis, $|f'(z)| \leqq MR$ on the real axis.*

Our generalization is an extension to functions of exponential type of A. Zygmund's integral analogue of Lemma 4.5 for trigonometrical polynomials.[22]

---

[18] [8], vol. 1, section III, problem 165, pp. 116, 289.
[19] [8], vol. 1, section III, problem 322, p. 145.
[20] [4], p. 417; see also [2]. Lemma 4.4 is a very special case of a general theorem.
[21] [1], p. 102; [8], vol. 2, section IV, problem 201, pp. 35, 218.
[22] [13], p. 394.

LEMMA 4.6.[23] *If $f(z)$ is of type R, if $\varphi(t)$ is non-negative, strictly increasing, and convex for $t \geqq 0$, and if*

$$\int_{-\infty}^{\infty} \varphi(|f(x)|)\,dx \tag{4.6}$$

if finite, then

$$\int_{-\infty}^{\infty} \varphi(|f'(x)|/R)\,dx \leqq \int_{-\infty}^{\infty} \varphi(|f(x)|)\,dx. \tag{4.7}$$

COROLLARY 4.6.1. *If $f(z)$ is of type R and belongs to L^p $(p \geqq 1)$ on the real axis, $f'(z)$ belongs to L^p on the real axis.*

We consider the function

$$g(z) = \int_0^z f(\zeta)\,d\zeta.$$

It is easy to verify that $g(z)$ is of type R. For real $x > 0$ we have, taking the line of integration to be the segment $(0, x)$, and applying Jensen's inequality,[24]

$$\varphi\left(x^{-1}\int_0^x |f(t)|\,dt\right) \leqq x^{-1}\int_0^x \varphi(|f(t)|)\,dt,$$

$$|g(x)| \leqq \int_0^x |f(t)|\,dt \leqq x\varphi^{-1}\left\{x^{-1}\int_0^x \varphi(|f(t)|)\,dt\right\}$$

$$\leqq Ax,$$

where A is a constant; similar reasoning with $x < 0$ shows that

$$|g(x)| \leqq A|x| \qquad (-\infty < x < \infty). \tag{4.8}$$

Let t be any real number, and write (to simplify the notation) $\rho = \pi/(2R)$. Then, because of (4.8), the function $h(z) = g(z + t - \rho)$ satisfies the hypotheses of Lemma 4.3. Therefore (4.3) is true for $h(x)$; that is, for real x,

$$\frac{d^2}{dx^2}\frac{g(x+t-\rho)}{\sin Rx} = 2R^2\sum_{n=-\infty}^{\infty}\frac{(-1)^n g(t-\rho+n\pi/R)}{(Rx-n\pi)^3}; \tag{4.9}$$

since $g'(x) = f(x)$, (4.9), with the differentiation carried out, gives us

$$\left|\frac{f'(x+t-\rho)}{\sin Rx}\right| = \left|\frac{2Rf(x+t-\rho)\cos Rx}{\sin^2 Rx} - R^2 g(x+t-\rho)\frac{1+\cos^2 Rx}{\sin^3 Rx}\right.$$

$$\left. + 2R^2\sum_{n=-\infty}^{\infty}\frac{(-1)^n g(t-\rho+n\pi/R)}{(Rx-n\pi)^3}\right|.$$

---

[23] This result was stated by the author, [3], p. 303, with a somewhat inadequate indication of a proof.

[24] [5], p. 152.

Now let $x = \rho$. Then, applying (4.8), we find

$$|f'(t)| \leqq R^2|g(t)| + \frac{16R^2}{\pi^3}\sum_{n=-\infty}^{\infty}\frac{|g(t+\rho-n\pi/R)|}{|1-2n|^3}$$

$$\leqq AR^2\left\{|t| + \frac{16}{\pi^3}\sum_{n=-\infty}^{\infty}\frac{(2|n|+1)\pi/(2R)+|t|}{|1-2n|^3}\right\}$$

$$< B(1+|t|),$$

where B is some constant. We now have, as $|x| \to \infty$, $g(x) = O(|x|)$ and $g''(x) = O(|x|)$; by Lemma 4.4, $f(x) = g'(x) = O(|x|)$. We can now apply Lemma 4.3 to $f(x)$; (4.2) gives us, for all real t, with $S > R$ and $\sigma = \pi/(2S)$,

$$|f'(t)| \leqq \frac{4S}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{|f(t+\sigma-n\pi/S)|}{(1-2n)^2},$$

provided that the series on the right converges. Applying Jensen's inequality,[25] we obtain

$$\varphi(|f'(t)|/S) \leqq \frac{4}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{\varphi(|f(t+\sigma-n\pi/S)|)}{(1-2n)^2}, \tag{4.10}$$

again if the series on the right converges.

We have assumed that (4.6) is finite; hence

$$\int_{-\infty}^{\infty}\varphi(|f(t-\sigma+n\pi/S)|)\,dt$$

is finite for each integer n, and equal to (4.6). Hence, since

$$\frac{4}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{1}{(1-2n)^2} = 1,$$

the series

$$\frac{4}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{1}{(1-2n)^2}\int_{-\infty}^{\infty}\varphi(|f(t-\sigma+n\pi/S)|)\,dt \tag{4.11}$$

converges, and its sum is (4.6). Since φ is non-negative, we may interchange summation and integration in (4.11), so that

$$\int_{-\infty}^{\infty}\varphi(|f(t)|)\,dt = \int_{-\infty}^{\infty}\frac{4}{\pi^2}\sum_{n=-\infty}^{\infty}\frac{\varphi(|f(t-\sigma+n\pi/S)|)}{(1-2n)^2}\,dt.$$

Therefore the series under the integral sign on the right converges, and (4.10) is valid. Integration of (4.10) gives the relation

$$\int_{-\infty}^{\infty}\varphi(|f'(t)|/S)\,dt \leqq \int_{-\infty}^{\infty}\varphi(|f(t)|)\,dt; \tag{4.12}$$

---

[25] [5], p. 74.

this is valid for every $S > R$. If we let $S \to R$ through a decreasing sequence, $|f'(t)|/S$ is a non-decreasing sequence for each t, and the sequence $\varphi(|f'(t)|/S)$ is a non-decreasing sequence of continuous non-negative functions. We may therefore let $S \to R$ under the integral sign in (4.12), and the result is (4.7).

5. Functions of zero type

We digress to apply Lemma 4.6 to the theory of functions of zero type.

THEOREM 5.1. *If $f(z)$ is of zero type, $f(z) \not\equiv 0$, and $\varphi(t)$ is non-negative, strictly increasing, and convex for $t \geqq 0$, $\varphi(|f(z)|)$ cannot be integrable on the real axis.*

Thus, in particular, $f(z)$ cannot belong to any L^p ($p \geqq 1$) on the real axis. This result, for $p = 1$, was given by C. L. Siegel;[26] for $p = 2$, by Paley and Wiener; and for $1 < p < 2$, by M. Plancherel and G. Pólya.[27]

If $\varphi(|f(z)|)$ is integrable on the real axis, and $f(z) \not\equiv 0$, (4.8) is valid (by Lemma 4.6) for any $R > 0$. As $R \to 0$, $|f'(x)|/R$ increases and becomes infinite (except on the countable set of points where $f'(x) = 0$). Since $\varphi(t)$ is convex and increasing, $\varphi(t)$ increases and becomes infinite as $t \to \infty$. Hence $\varphi(|f'(x)|/R)$, as a function of R, increases and becomes infinite, as $R \to 0$, for almost all x. Therefore

$$\lim_{R \to 0} \int_{-\infty}^{\infty} \varphi(|f'(x)|/R)\, dx = \infty,$$

which contradicts (4.8).

If we specialize $\varphi(t)$, we can obtain more detailed results.

THEOREM 5.2. *If $f(z)$ is of zero type, and $f(x)/(1 + |x|^n)$ belongs to $L^p(-\infty, \infty)$ ($p \geqq 1$), for an integer $n \geqq 0$, then $f(z)$ is a polynomial of degree at most $n - 1$ ($n - 2$ if $p = 1$).*

The case $n = 0$ is covered by Theorem 5.1; we assume now that $n > 0$. We consider first the case $p > 1$. Then the function

$$g(z) = z^{-n}\left\{f(z) - \sum_{k=0}^{n-1} \frac{z^k f^{(k)}(0)}{k!}\right\}$$

is entire, of zero type, and belongs to L^p on the real axis; by Theorem 5.1, $g(z) \equiv 0$, and our conclusion follows.

If $p = 1$, we again form the function $g(z)$. If $g(z)$ has a zero a, the function

$$h(z) = (z-a)^{-1} g(z)$$

is entire, of zero type, and belongs to L^1 on the real axis. Then $h(z) \equiv 0$ and $g(z) \equiv 0$, and $f(z) = Bz^{n-1} + P_{n-2}(z)$, where $P_{n-2}(z)$ is a polynomial of degree at most $n - 2$. If, however, $g(z)$ were never zero, there would exist an entire function $h(z)$ such that $h(z)^2 = g(z)$; clearly, $h(z)$ would be of zero type; and $h(x)/(1 + |x|)$ would belong to $L^2(-\infty, \infty)$. By the first part of the proof, $h(z)$ would be a constant; therefore $g(z)$ would be a constant, and we should

[26] [9], p. 323.
[27] [7], p. 245.

have $f(z) = Az^n + Bz^{n-1} + P_{n-2}(z)$, where $P_{n-2}(z)$ would be a polynomial of degree at most $n - 2$. But if $[Ax^n + Bx^{n-1} + P_{n-2}(x)]/(1 + |x|^n)$ belongs to $L^1(-\infty, \infty)$, $A = B = 0$; thus in either case $f(z) = P_{n-2}(z)$ (and $g(z) \equiv 0$, so that only the first case can arise).

6. Functions of class $L^p(1 < p \leqq 2)$

We shall need a weak form of the Hausdorff-Young theorem, and the Fourier transform analogue of the other half; we state these results as lemmas.

LEMMA 6.1.[28] *If the complex numbers c_n $(n = 0, \pm 1, \pm 2, \cdots)$ satisfy*

$$\sum_{n=-\infty}^{\infty} |c_n|^p < \infty \qquad (1 < p \leqq 2), \tag{6.1}$$

there exists, for any $R > 0$, a function $a(t)$ of $L^{p/(p-1)}(-R, R)$, such that

$$c_n = \frac{1}{2R} \int_{-R}^{R} e^{-in\pi t/R} a(t)\, dt. \tag{6.2}$$

LEMMA 6.2.[29] *If the function $g(t)$ is of class $L^p(-\infty, \infty)$ $(1 < p \leqq 2)$, then*

$$f(x) = \operatorname*{l.i.m.}_{T\to\infty}^{(q)} \int_{-T}^{T} e^{ixt} g(t)\, dt \qquad (q = p/(p-1)) \tag{6.3}$$

is defined for almost all x, and belongs to $L^q(-\infty, \infty)$.

We shall also need a consequence of Corollary 4.6.1.

LEMMA 6.3. *If $f(z)$ is of type R and belongs to L^p $(p \geqq 1)$ on the real axis, the series*

$$\sum_{n=-\infty}^{\infty} |f(n\pi/R)|^p \tag{6.4}$$

converges.[29a]

We suppose that $f(x)$ is real; in the general case, we should break $f(x)$ into its real and imaginary parts, and apply Minkowski's inequality.

Since

$$\sum_{n=-\infty}^{\infty} \int_{n\pi/R}^{(n+1)\pi/R} |f(t)|^p\, dt = \int_{-\infty}^{\infty} |f(t)|^p\, dt < \infty,$$

(6.4) will converge if

$$\sum_{n=-\infty}^{\infty} \left| \left|\frac{\pi}{R}\right| f\left(\frac{n\pi}{R}\right)\right|^p - \int_{n\pi/R}^{(n+1)\pi/R} |f(t)|^p\, dt \Bigg| \tag{6.5}$$
$$= \sum_{n=-\infty}^{\infty} \left| \int_{n\pi/R}^{(n+1)\pi/R} \left\{ \left| f\left(\frac{n\pi}{R}\right)\right|^p - |f(t)|^p \right\} dt \right|$$

---

[28] [14], p. 190.

[29] [14], p. 316.

[29a] A more general result is given by Plancherel and Pólya in the second part of [7], which appeared while this paper was in press.

converges. But $|f(t)|^p$ is absolutely continuous, so that for $n\pi/R \leqq t \leqq (n+1)\pi/R$,

$$\left|\left|f\left(\frac{n\pi}{R}\right)\right|^p - |f(t)|^p\right| \leqq \int_{n\pi/R}^{(n+1)\pi/R} \left|\frac{d}{dt}|f(t)|^p\right| dt$$

$$\leqq p\int_{n\pi/R}^{(n+1)\pi/R} |f(t)|^{p-1}|f'(t)|\,dt.$$

Hence (6.4) converges if

$$\int_{-\infty}^{\infty} |f(t)|^{p-1}|f'(t)|\,dt < \infty. \tag{6.6}$$

Relation (6.6) is the conclusion of Corollary 4.6.1 if $p = 1$; if $p > 1$, we have by Hölder's inequality

$$\int_{-\infty}^{\infty} |f(t)|^{p-1}|f'(t)|\,dt \leqq \left(\int_{-\infty}^{\infty} |f'(t)|^p\,dt\right)^{(p-1)/p}\left(\int_{-\infty}^{\infty} |f(t)|^p\,dt\right)^{1/p},$$

and the right hand side is finite by Corollary 4.6.1. This establishes the lemma.

We can now establish

THEOREM 6.4.[30] *If $f(z)$ is of type R and belongs to L^p $(1 < p \leqq 2)$ on the real axis, then*

$$f(z) = \int_{-R}^{R} e^{izt}a(t)\,dt, \tag{6.7}$$

where $a(t)$ belongs to $L^{p/(p-1)}(-R, R)$.

We shall apply Theorem 3.1, with p replaced by $p/(p-1)$. We set

$$c_n = \frac{1}{2R}f\left(\frac{-n\pi}{R}\right) \qquad (n = 0, \pm1, \pm2, \cdots).$$

By Lemma 6.3, the series (6.1) converges; by Lemma 6.1, there is a function $a(t)$ of $L^{p/(p-1)}(-R, R)$, such that

$$\frac{1}{2R}f\left(\frac{-n\pi}{R}\right) = \frac{1}{2R}\int_{-R}^{R} e^{-in\pi t/R}a(t)\,dt. \tag{6.8}$$

Relation (6.8) is the same as (3.4); if we can establish (3.2) we shall be able to apply Theorem 3.1.

Let E denote the set of points x for which $|f(x)| > x^{-1/(2p)}$. Then

$$A = \int_{-\infty}^{\infty} |f(t)|^p\,dt \geqq \int_E |f(t)|^p\,dt > \int_E t^{-\frac{1}{2}}\,dt. \tag{6.9}$$

If $0 < \delta < \pi/(2R)$, an infinite number of the intervals

$$I_n: \left(\frac{n\pi}{2R} - \delta, \quad \frac{n\pi}{2R}\right) \qquad (n = \pm1, \pm2, \cdots)$$

[30] Essentially contained in [7], p. 243.

must contain points of E', the complement of E. In fact, if this were not true, E would contain every I_n for n greater than some n_0. We should then have, because of (6.9),

$$\sum_{|n|>n_0} \frac{\delta}{2}\left(\frac{n\pi}{2R}\right)^{-\frac{1}{2}} < \sum_{|n|>n_0} \frac{\frac{n\pi}{2R} - \left(\frac{n\pi}{2R} - \delta\right)}{\left(\frac{n\pi}{2R}\right)^{\frac{1}{2}} + \left(\frac{n\pi}{2R} - \delta\right)^{\frac{1}{2}}}$$

$$= \tfrac{1}{2} \sum_{|n|>n_0} \int_{n\pi/(2R)-\delta}^{n\pi/(2R)} t^{-\frac{1}{2}}\, dt < A,$$

which is evidently impossible. We can therefore find a sequence $\{x_{n_k}\}$ with $n_k\pi/(2R) - \delta < x_{n_k} < n_k\pi/(2R)$, such that

$$|f(x_{n_k})| < x_{n_k}^{-1/(2p)};$$

and since

$$|x_{n_k} - n_k\pi/R| > \pi/(2R) - \delta,$$

(3.2) is satisfied. We now apply Theorem 3.1 and obtain our conclusion.

As a counterpart to Theorem 6.4, we establish

THEOREM 6.5. *If*

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt, \tag{6.10}$$

with $a(t)$ of class $L^p(-R, R)$ $(1 < p \leqq 2)$, then $f(z)$ is of type R and belongs to $L^{p/(p-1)}$ on the real axis.

When $p = 2$, $p/(p-1) = 2$, and Theorems 6.4 and 6.5 coalesce into the theorem of Paley and Wiener cited in §1.

Theorem 6.5 is an immediate consequence of Lemma 6.2, since under the hypotheses of Theorem 6.5, the limit in mean of Lemma 6.2 reduces to the ordinary limit of a sequence whose terms are ultimately all the same.

7. Examples

As one would expect from the theory of Fourier series, Theorems 6.4 and 6.5 fail when $p > 2$.

THEOREM 7.1. *For any $p > 2$, there exists a function $f(z)$ of type R, belonging to L^p on the real axis, but not of the form* (6.7) *with $a(t)$ of $L^{p/(p-1)}(-R, R)$.*

Take constants α, β, with $0 < \alpha < \beta < 1$, $\alpha < 1 - 2p^{-1}$, and $\beta < \alpha p/(p-1)$; define an odd function $b(t)$ by[31]

$$b(t) = n^\alpha \cos(n^2\pi t/R),\ R(n+1)^{-\beta} < t \leqq Rn^{-\beta} \qquad (n = 1, 2, \cdots),$$

$$b(t) = 0,\ t > R.$$

[31] This example is slightly modified from one given by Titchmarsh, [10], pp. 441–442. Titchmarsh sketches, for $x = n\pi/R$, a proof of (7.1) which is readily extended to general x.

It can then be shown that $|b(t)|$ is integrable on $(-R, R)$, while $|b(t)|^{p/(p-1)}$ is not; and that, if

$$(z) = \int_{-R}^{R} e^{izt} b(t)\, dt,$$

$$f(x) = O(|x|^{\frac{1}{2}\alpha-\frac{1}{2}} \log|x|) = O(|x|^{-1/p-\epsilon} \log|x|) \qquad (|x| \to \infty), \tag{7.1}$$

where $\epsilon = \frac{1}{2}(1 - 2p^{-1} - \alpha)$, so that $f(x)$ belongs to $L^p(-\infty, \infty)$. Furthermore, $f(z)$ is of type R, by Lemma 2.1.

THEOREM 7.2. *For any $p > 2$, there exists a function $a(t)$ of class $L^p(-R, R)$, such that*

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt$$

does not belong to $L^{p/(p-1)}$ on the real axis.

Let $a(t)$ be a continuous function on $(-R, R)$, such that if

$$c_n = \frac{1}{2R} \int_{-R}^{R} e^{in\pi t/R} a(t)\, dt,$$

the series

$$\sum_{n=-\infty}^{\infty} |c_n|^r \tag{7.2}$$

diverges for every r, $1 < r < 2$.[32] Form

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt.$$

Then $f(z)$ cannot belong to $L^{p/(p-1)}$ on the real axis, since if it did, (7.2) would converge with $r = p/(p-1) < 2$, by Lemma 6.3.

8. Functions of L^p for $p = 1$ and $p > 2$

It is natural to ask what sort of integral representation can be obtained for a function of type R which belongs on the real axis to a class L^p where p does not satisfy $1 < p \leqq 2$. To answer this question for $p = 1$, we shall use the following theorem of N. Wiener.

LEMMA 8.1.[33] *Let $a(t)$ be continuous on $(-\infty, \infty)$, and zero outside $(-R, R)$. Let*

$$f(x) = \int_{-R}^{R} e^{izt} a(t)\, dt.$$

Then

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty \tag{8.1}$$

---

[32] [14], p. 119.

[33] [12], p. 80.

if and only if for every $S > R$

$$\sum_{n=-\infty}^{\infty} |f(n\pi/S)| < \infty. \tag{8.2}$$

We remark that our Lemma 6.3 furnishes a proof that (8.1) implies (8.2), and even that

$$\sum_{n=-\infty}^{\infty} |f(n\pi/R)| < \infty. \tag{8.3}$$

THEOREM 8.2.[34] *A necessary and sufficient condition that* $f(z)$ *be of type* R *and belong to* L^1 *on the real axis is that* $f(z)$ *have the form*

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt, \tag{8.4}$$

where $a(R-) = a(-R+) = 0$ *and* $a(t)$ *has an absolutely convergent Fourier series over* $(-R, R)$.

Suppose that $f(z)$ is of type R and belongs to L^1 on the real axis. The proof of Theorem 3.1 shows that $f(z)$ will have the form (8.4) with $a(t)$ continuous on $(-R, R)$ if $f(x)$ satisfies (3.2) for a sequence satisfying (3.3), and the condition

$$f(n\pi/R) = \int_{-R}^{R} e^{-in\pi t/R} a(t)\, dt, \qquad a(t) \text{ continuous.} \tag{8.5}$$

Since (8.3) converges,

$$\sum_{n=-\infty}^{\infty} \frac{1}{2R} f(-n\pi/R)\, e^{in\pi t/R}$$

is absolutely convergent, and is therefore the Fourier series of a continuous function $a(t)$ which satisfies (8.5). Condition (3.2) is established as in the proof of Theorem 6.4. Hence $f(z)$ has the form (8.4), and it remains only to show that $a(R-) = a(-R+) = 0$. To do this, we note that $f(z)$ is of type S for $S > R$, and hence that for some $S > R$

$$f(z) = \int_{-S}^{S} e^{izt} b(t)\, dt$$

where $b(t)$ is continuous on $(-S, S)$. But

$$f(z) = \int_{-R}^{R} e^{izt} a(t)\, dt = \int_{-S}^{S} e^{izt} c(t)\, dt,$$

where

$$c(t) = \begin{cases} a(t), & |t| < R. \\ 0, & |t| > R, \end{cases}$$

[34] Compare [7], pp. 229 ff.

Hence $b(t) \equiv c(t)$ $(-S < t < S)$, and since $b(t)$ is continuous at $t = \pm R$, so is $c(t)$, and $a(R-) = a(-R+) = 0$.

If $f(z)$ has the form (8.4) with $a(t)$ satisfying the specified conditions, it is clear that (8.2) is satisfied, since

$$\sum_{n=-\infty}^{\infty} \frac{1}{2S} f(-n\pi/S)\, e^{in\pi t/S}$$

is the Fourier series on $(-S, S)$ of the function $b(t)$, equal to $a(t)$ on $(-R, R)$, and zero outside $(-R, R)$. By Lemma 8.1, (8.1) is satisfied; that is, $f(z)$ belongs to L^1 on the real axis.

Finally, we establish

THEOREM 8.3. *If $f(z)$ is of type R, and of class L^p $(p > 2)$ on the real axis,*

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t), \tag{8.6}$$

where $\alpha(t)$ is the integral[35] *of order δ of a function of class $L^2(-R, R)$ for every δ less than $\frac{1}{2} + 1/p$ (and is consequently continuous on $-R \leqq t \leqq R$).*

By a theorem of Hardy and Littlewood,[36] $\alpha(t)$ belongs to the Lipschitz class Lip η for $0 \leqq \eta < p^{-1}$. It is interesting to compare this property of $\alpha(t)$ with a corresponding property when $1 < p \leqq 2$. By Theorem 6.4, if $f(z)$ is of type R and of class L^p $(1 < p \leqq 2)$ on the real axis, $f(z)$ has the form (8.6) with $\alpha(t)$ the integral of a function of class $L^{p/(p-1)}(-R, R)$; $\alpha(t)$ consequently[37] belongs to the class Lip p^{-1}. When $p = 1$, $f(z)$ has (by Theorem 8.2) the form (8.6) with $\alpha(t)$ the integral of a bounded function, and consequently of class Lip 1. On the other hand, in the limiting case $p = \infty$, $f(z)$ can have the form (8.6) with $\alpha(t)$ not of class Lip 0 (i.e., with $\alpha(t)$ discontinuous); trigonometric polynomials furnish examples.

To establish Theorem 8.3, we apply Theorem 3.2. Let S be any number greater than R. By Lemma 6.3, the series

$$\sum_{n=-\infty}^{\infty} |f(n\pi/S)|^p$$

converges. By Hölder's inequality,

$$\sum_{n=-\infty}^{\infty}{}' |f(n\pi/S)|\, n^{-1} \leqq \left(\sum_{n=-\infty}^{\infty}{}' |f(n\pi/S)|^p\right)^{1/p} \left(\sum_{n=-\infty}^{\infty} |n|^{-p/(p-1)}\right)^{(p-1)/p},$$

so that the series on the left converges, and

$$\sum_{n=-\infty}^{\infty}{}' \frac{1}{2in\pi} f\left(\frac{-n\pi}{S}\right) e^{in\pi t/S}$$

---

[35] In the sense of H. Weyl; see [14], p. 223.

[36] [14], p. 225.

[37] [14], p. 225.

is the Fourier series of a function $\gamma(t)$, continuous on $-S \leqq t \leqq S$. By Theorem 3.2,

$$f(z) = \int_{-S}^{S} e^{izt}\, d\alpha_S(t),$$

where $\alpha_S(t)$ is continuous on $-S < t < S$. We may suppose that $\alpha_S(0) = 0$. Then, since

$$\int_{-T}^{T} e^{izt}\, d\alpha(t) = \int_{-T}^{T} e^{izt}\, d\beta(t),$$

with $\alpha(t)$ and $\beta(t)$ continuous on $-T < t < T$ and $\alpha(0) = \beta(0)$, implies $\alpha(t) = \beta(t)$ for $-T < t < T$, it follows that all $\alpha_S(t)$ are equal to a continuous function $\alpha(t)$ on $-R < t < R$, and equal to $\alpha(-R+)$ or $\alpha(R-)$ on $(-S, -R)$ and (R, S), respectively. Hence

$$f(z) = \int_{-R}^{R} e^{izt}\, d\alpha(t),$$

with $\alpha(t)$ continuous on $-R \leqq t \leqq R$.

If $\sigma > (p-2)/(2p)$, we have, for $S > R$,

$$\sum_{n=-\infty}^{\infty}{}' |f(n\pi/S)|^2 |n|^{-2\sigma} \leqq \left(\sum_{n=-\infty}^{\infty}{}' |f(n\pi/S)|^p\right)^{2/p} \left(\sum_{n=-\infty}^{\infty}{}' |n|^{-2\sigma p/(p-2)}\right)^{(p-2)/p},$$

so that the series on the left converges, and

$$\sum_{n=-\infty}^{\infty}{}' \frac{f(-n\pi/S)}{(2in\pi)^\sigma}\, e^{in\pi t/S}$$

is the Fourier series of a function $\beta(t)$ belonging to $L^2(-S, S)$. The integral of order $(1-\sigma)$ of $\beta(t)$ has the Fourier series[38]

$$\sum_{n=-\infty}^{\infty}{}' \frac{f(-n\pi/S)}{2in\pi}\, e^{in\pi t/S},$$

and hence is identically equal to $\gamma(t)$. Since $\alpha(t)$ is obtained from $\gamma(t)$ by adding a function of class C', $\alpha(t)$ is also the integral of order $(1-\sigma)$ of a function of L^2. If $\sigma > (p-2)/(2p)$, $\delta = 1 - \sigma < (p+2)/(2p)$.

REFERENCES

1. S. Bernstein, *Leçons sur les propriétés extrêmales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris, 1926.
2. R. P. Boas, Jr., *Asymptotic relations for derivatives*, Duke Mathematical Journal, 3 (1937), 637–646.
3. R. P. Boas, Jr., *Some theorems on Fourier transforms and conjugate trigonometric integrals*, Transactions of the American Mathematical Society, 40 (1936), 287–308.
4. G. H. Hardy and J. E. Littlewood, *Contributions to the arithmetic theory of series*, Proceedings of the London Mathematical Society, (2), 11 (1913), 411–478.

[38] [14], p. 223.

5. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1934.
6. R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*, New York, 1934.
7. M. Plancherel and G. Pólya, *Fonctions entières et intégrales de Fourier multiples*, Commentarii Mathematici Helvetici, 9 (1936–37), 224–248; 10 (1937–38), 110–163.
8. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925.
9. C. L. Siegel, *Über Gitterpunkte in convexen Körpern und ein damit zusammenhängendes Extremalproblem*, Acta Mathematica, 65 (1935), 307–323.
10. E. C. Titchmarsh, *The theory of functions*, Oxford, 1932.
11. G. Valiron, *Sur la formule d'interpolation de Lagrange*, Bulletin des Sciences Mathématiques, (2), 39 (1925), 181–192, 203–224.
12. N. Wiener, *The Fourier integral and certain of its applications*, Cambridge, 1933.
13. A. Zygmund, *A remark on conjugate series*, Proceedings of the London Mathematical Society, (2), 34 (1932), 392–400.
14. A. Zygmund, *Trigonometrical series*, Warsaw-Lwów, 1935.
15. Loo-Keng Hua and Shien-siu Shü, *On Fourier transforms in L^p in the complex domain*, Journal of Mathematics and Physics, 15 (1936), 249–263.
16. A. C. Offord, *On Fourier transforms*. III., Transactions of the American Mathematical Society, 38 (1935), 250–266.

Princeton University.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

CLOSURE THEOREMS FOR TRANSLATIONS

BY R. P. BOAS, JR.,[1] AND S. BOCHNER

(Received January 20, 1937)

Introduction

If $f(t)$ is a function defined for $-\infty < t < \infty$, we call $f(t+u)$ a translation of $f(t)$. The closure of the set of all translations of a given function in the spaces $L(-\infty, \infty)$ and $L^2(-\infty, \infty)$ has been investigated by N. Wiener,[2] and he has characterized the functions whose translations are dense in these spaces as those whose Fourier transforms are different from zero (everywhere, for L; almost everywhere, for L^2). We shall discuss the closure in certain spaces of the translations of functions which do not necessarily have Fourier transforms, but do have generalized Fourier transforms of the kind defined by S. Bochner.[3] Since the spaces which we consider are adapted to the least upper bound of the difference as metric, our results do not overlap Wiener's, but may be considered to be dual to them.

Let $p(t)$ be a positive continuous function, decreasing as $t \to \pm\infty$, such that $p(t)/p(2t)$ is uniformly bounded and $|t|^{-n} = O(p(t))$ $(|t| \to \infty)$ for some integer n $(n \geq 0)$. Then the set $D_{p(t)}$ of all complex-valued functions $g(t)$, continuous except for discontinuities of the first kind, such that $\lim_{t\to\pm\infty} p(t)g(t)$ exists, can be metrised by means of the norm $\|g(t)\| = \max_{-\infty\leq t\leq\infty} p(t)|g(t)|$; and every such function $g(t)$ has a generalized Fourier transform of order $n+2$, $\gamma(t)$. We shall show that the following statements are equivalent.

(a) *There exists no interval in which $\gamma(t)$ is equal to a polynomial of degree at most $n+1$.*

(b) *For any continuous function $f(t)$ with $\lim_{t\to\pm\infty} p(t)f(t) = 0$, and for any $\epsilon > 0$, there exist real numbers λ_j, complex numbers A_j, and an integer k, such that*

$$\max_{-\infty\leq t\leq\infty} p(t)\left|f(t) - \sum_{j=1}^{k} A_j g(t+\lambda_j)\right| < \epsilon. \tag{1}$$

Statement (b) states that the translations of $g(t)$ are closed in the set of all continuous functions $f(t)$ of the space, for which $\lim_{t\to\pm\infty} p(t)f(t) = 0$. Our theorem is thus an unsymmetric approximation theorem, in which the approximating functions, while drawn from a restricted class, are more irregular in

[1] National Research Fellow.

[2] N. Wiener, *The Fourier integral and certain of its applications*, Cambridge, 1933, pp. 97–103. This book will be cited as "*W*".

[3] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932, Chapter 6. This book will be cited as "*B*".

their behavior than the functions approximated. This state of affairs is to be expected, since, for example, a discontinuous function could not be approximated in the sense of relation (1) by continuous functions. It is even true that the theorem becomes more interesting as the approximating functions become more general, since approximation to a regular function $f(t)$ by irregular functions $g(t + \lambda_j)$ imposes restrictions on the coefficients A_j in (b). A peculiar feature of our theorem is that alteration of $g(t)$ on an isolated set of points does not affect the Fourier transform of $g(t)$, and hence does not affect the ability of the translations of $g(t)$ to approximate to continuous functions. However, such an alteration will in general affect the numbers λ_j and A_j of (b), since the approximation which we obtain is uniformly close over every finite interval (not merely approximation on the average, as in Wiener's theorem).

We base our proof on the well known equivalence, established by H. Hahn, between closure and completeness in a normed vectorial space;[4] this equivalence subsists for the complex spaces which we consider. In order to apply it, it is necessary to know a general representation for linear functionals on the space $D_{p(t)}$; this is obtained easily from the corresponding representation for the space of functions, on a finite interval, with at most discontinuities of the first kind—a representation which has been obtained by H. S. Kaltenborn.[5] We also use important theorems of N. Wiener, which he has used in discussing the closure of translations of functions of L.

We also consider the simpler problems of the closure of translations of periodic functions which have only discontinuities of the first kind, and of the closure of the translations of a sequence of numbers.

1. Lemmas on functionals

We shall use the terminology of Banach's book.[6] We shall have to consider, in addition to the ordinary vectorial spaces, in which multiplication of elements by real numbers is defined, vectorial spaces in which multiplication of elements by complex numbers is defined; we shall call them real vectorial spaces and complex vectorial spaces, respectively.

DEFINITION 1.1. *D shall denote the complex normed vectorial space whose elements are complex-valued functions $x(t)$ defined for $-1 \leqq t \leqq 1$, continuous except for discontinuities of the first kind. Addition of elements and multiplication of elements by constants are defined by the corresponding processes on the functions. The norm in D is $\| x(t) \| = \max_{-1 \leqq t \leqq 1} | x(t) |$.*

LEMMA 1.1 [H. S. Kaltenborn].[7] *Every linear functional $f(x)$ on D can be written in the form*

$$f(x) = \mathfrak{BD} \int_{-1}^{1} x(t)\, d\alpha(t) + \sum_{k=-\infty}^{\infty} \varphi(c_k)[x(c_k) - x(c_k-)], \tag{1.1}$$

[4] See S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932, p. 58.

[5] H. S. Kaltenborn, *Linear functional operations on functions having discontinuities of the first kind*, Bulletin of the American Mathematical Society, 40 (1934), 702–708.

[6] S. Banach, op. cit. In particular, "linear" means "additive and continuous."

[7] H. S. Kaltenborn, op. cit.

where $\alpha(t)$ is a function of bounded variation on $(-1, 1)$;

$$\mathfrak{BD}\int_a^b x(t)\,d\alpha(t) = A$$

if for every $\epsilon > 0$ there exist an integer n and points θ_i, $a = \theta_0 < \theta_1 < \cdots < \theta_{n+1} = b$, such that

$$\left|A - \sum_{i=0}^{m} x(\tau_i)[\alpha(t_{i+1}) - \alpha(t_i)]\right| < \epsilon \tag{1.2}$$

whenever

$$a = t_0 < t_1 < \cdots < t_{m+1} = b, \qquad t_i < \tau_i < t_{i+1} \qquad (i = 0, 1, \cdots, m),$$

and every θ_i $(i = 1, 2, \cdots, n)$ is equal to some t_j; $\phi(t)$ vanishes except on the countable set of points $\{c_k\}$, and has bounded variation.

The integral which we denote by $\mathfrak{BD}\int_a^b x(t)\,d\alpha(t)$ was introduced by B. Dushnik,[8] and he has shown that it exists when $x(t)$ has at most discontinuities of the first kind and $\alpha(t)$ is of bounded variation.

Lemma 1.1 is given by Kaltenborn for the real vectorial space consisting of the real functions in D. The extension to the whole of D is immediate.[9]

DEFINITION 1.2. *$D_{p(t)}$ shall denote the complex normed vectorial space whose elements are complex valued functions $x(t)$ defined on $(-\infty, \infty)$, continuous except for discontinuities of the first kind, and such that the limits $\lim_{t\to\pm\infty} p(t)x(t)$ exist, where $p(t)$ is positive, continuous, non-decreasing on $(-\infty, 0)$ and non-increasing on $(0, \infty)$; addition of elements and multiplication by constants are defined in the usual way; the norm in $D_{p(t)}$ is $\| x(t) \|_{p(t)} = \max_{-\infty\le t\le\infty} p(t)\,|\,x(t)\,|$.*

LEMMA 1.2. *Every linear functional on $D_{p(t)}$ can be written in the form*

$$f(x) = \mathfrak{BD}\int_{-\infty}^{\infty} x(t)\,d\alpha(t) + \sum_{k=-\infty}^{\infty} \varphi(c_k)[x(c_k) - x(c_k-)] + AL_+(x) + BL_-(x), \tag{1.3}$$

where

$$\int_{-\infty}^{\infty} p(t)^{-1}\,|\,d\alpha(t)\,| < \infty; \tag{1.4}$$

$\varphi(t)$ vanishes except on the countable set $\{c_k\}$, and

$$\sum_{k=-\infty}^{\infty} |\,\varphi(c_k)\,|/p(c_k) < \infty; \tag{1.5}$$

$$L_+(x) = \lim_{t\to+\infty} p(t)x(t), \qquad L_-(x) = \lim_{t\to-\infty} p(t)x(t); \tag{1.6}$$

and A and B are constants independent of $x(t)$.

---

[8] B. Dushnik, *On the Stieltjes integral*, Ann Arbor [?1931] (Dissertation, University of Michigan), pp. 14, 18.

[9] If $f(x)$ is linear on D, $f(x_1 + ix_2) = f(x_1) + if(x_2)$ (x_1 and x_2 real), and f is clearly linear on the space of real functions in D. For this space, $f(x) = f_1(x) + if_2(x)$, where f_1 and f_2 are real-valued functionals, clearly linear; each then has the form in question, and we obtain (1.1) by combining them.

We establish the lemma first for the space D_1 ($D_{p(t)}$ with $p(t) = 1$). Let $r(t)$ be a real-valued function, continuous and monotonic on $(-\infty, \infty)$, with $r(-\infty) = -1$, $r(\infty) = 1$ (e.g., $r(t) = (2/\pi)\tan^{-1} t$) The transformation which carries $y(t)$ into $z(t) = y(r(t))$ is a one-to-one and isometric transformation from D_1 to D; to a linear functional $g(y)$ on D_1 corresponds a linear functional $h(z) = g[y(r(t))]$ on D, and conversely. By Lemma 1.1,

$$h(z) = \mathfrak{BD} \int_{-1}^{1} z(t)\, d\gamma(t) + \sum_{k=-\infty}^{\infty} \chi(d_k)[z(d_k) - z(d_k-)],$$

where $\gamma(t)$ has bounded variation on $(-1, 1)$, and $\sum_{k=-\infty}^{\infty} |\chi(d_k)| < \infty$. This relation may conveniently be written

$$h(z) = \mathfrak{BD} \int_{-1}^{1} z(t)\, d\gamma_1(t) + \sum_{k=-\infty}^{\infty} \chi(d_k)[z(d_k) - z(d_k-)] + Az(1-) + Bz(-1+),$$

where $\gamma_1(1) = \gamma_1(1-)$, $\gamma_1(-1) = \gamma_1(-1+)$, $A = \gamma(1) - \gamma(1-)$, and $B = \gamma(-1+) - \gamma(-1)$. If now $t = r^{-1}(u)$, and $y(u) = z[r^{-1}(u)]$,

$$\begin{aligned} g(y) = h(z) = \mathfrak{BD} &\int_{-\infty}^{\infty} y(u)\, d\beta(u) \\ &+ \sum_{k=-\infty}^{\infty} \psi(c_k)[y(c_k) - y(c_k-)] + Ay(\infty) + By(-\infty), \end{aligned} \tag{1.7}$$

where $\beta(u) = \gamma_1[r^{-1}(u)]$, $c_k = r^{-1}(d_k)$, and $\psi(t) = \chi[r^{-1}(t)]$. Evidently

$$\int_{-\infty}^{\infty} |d\beta(u)| < \infty, \qquad \sum_{k=-\infty}^{\infty} |\psi(c_k)| < \infty.$$

This establishes Lemma 1.2 for the space D_1.

The transformation carrying $x(t)$ into $y(t) = x(t)p(t)$ is a one-to-one isometric transformation of $D_{p(t)}$ into D_1; it and its inverse preserve linear functionals. If $f(x)$ is a linear functional on $D_{p(t)}$, set $f(x) = g(y)$; then $g(y)$ has the form (1.7). Set

$$\alpha(t) = \int_{-\infty}^{t} p(u)\, d\beta(u), \qquad \varphi(t) = p(t)\psi(t);$$

then (1.7) becomes (1.3), and (1.5) is obvious; if $V(t)$ is the total variation of $\alpha(u)$ on $-\infty < u \leqq t$, we have

$$V(t) = \int_{-\infty}^{t} p(u)\, |d\beta(u)|,$$

$$\int_{-\infty}^{\infty} p(t)^{-1}\, dV(t) = \int_{-\infty}^{\infty} |d\beta(u)| < \infty,$$

so that (1.4) is satisfied.

LEMMA 1.3. *If $p(t)$ has the same properties as in Definition* 1.2, *and if $\alpha(t)$ is a function of bounded variation satisfying* (1.4), *then*

$$\int_{-\infty}^{\infty} p(t)^{-1} \,|\, \alpha(t+h) - \alpha(t) \,|\, dt < \infty. \tag{1.8}$$

We suppose, for definiteness, that $h > 0$, and consider convergence of the integral at the upper limit of integration. Let $V(t)$ be the total variation of $\alpha(u)$ for $0 \leqq u \leqq t$; (1.4) implies that

$$\beta(t) = -\int_t^{\infty} p(u)^{-1}\, dV(u)$$

approaches zero as $t \to \infty$. For $R' > 0$, $R'' > 0$,

$$\begin{aligned}\int_{R'}^{R''} |\alpha(t+h) - \alpha(t)|\, p(t)^{-1}\, dt &\leqq \int_{R'}^{R''} |V(t+h) - V(t)|\, p(t)^{-1}\, dt \\ &= \int_{R'}^{R''} p(t)^{-1}\, dt \int_t^{t+h} dV(u) = \int_{R'}^{R''} p(t)^{-1}\, dt \int_t^{t+h} p(u)\, d\beta(u) \\ &\leqq \int_{R'}^{R''} dt \int_t^{t+h} d\beta(u) = \left(\int_{R''}^{R''+h} - \int_{R'}^{R'+h}\right) \beta(t)\, dt,\end{aligned}$$

and the last expression clearly approaches zero as R' and R'' become infinite.

LEMMA 1.4. *If $x(t)$ belongs to $D_{p(t)}$, and $\alpha(t)$ satisfies* (1.4),

$$\int_h^{h+v} du\, \mathfrak{BD} \int_{-\infty}^{\infty} x(t)\, d\alpha(t+u) = \int_{-\infty}^{\infty} x(t)[\alpha(t+h+v) - \alpha(t+h)]\, dt. \tag{1.9}$$

It is clearly sufficient to consider real functions $x(t)$ and $\alpha(t)$, and positive h and v. By Lemma 1.3, the integral on the right of (1.9) is absolutely convergent. If we replace $x(t)$ by the function $x_n(t)$ which is equal to $x(t)$ on $(-n, n)$ and to zero outside $(-n, n)$, we shall approximate to both sides of (1.9) arbitrarily closely for large n. Hence it is sufficient to establish Lemma 1.4 for functions $x(t)$ vanishing outside a finite interval. But, as Kaltenborn has shown,[10] such a function can be uniformly approximated by finite linear combinations of functions of the forms

$$y(t) = \begin{cases} 0, & t \leqq a, \\ 1, & a < t < b, \\ 0, & t \geqq b, \end{cases} \qquad \text{and} \qquad z(t) = \begin{cases} 0, & t \neq a, \\ 1, & t = a. \end{cases}$$

It is therefore sufficient to establish Lemma 1.4 for functions $y(t)$ and $z(t)$, for which we can evaluate both sides of (1.9) explicitly. Referring to the definition of the $\mathfrak{BD}$ integral in Lemma 1.1, we see that both sides of (1.9) vanish if $z(t)$ is substituted for $x(t)$. To evaluate the left hand side of (1.9) with $x(t)$

[10] H. S. Kaltenborn, op. cit., p. 704.

replaced by $y(t)$, we take a and b as division points in the sum whose limit is the $\mathcal{BD}$ integral; we see that

$$\mathcal{BD}\int_{-\infty}^{\infty} y(t)\,d\alpha(t+u) = \alpha(b+u) - \alpha(a+u),$$

$$\int_h^{h+v} du\,\mathcal{BD}\int_{\infty}^{\infty} y(t)\,d\alpha(t+u) = \int_{b+h}^{b+h+v} \alpha(u)\,du - \int_{a+h}^{a+h+v} \alpha(u)\,du.$$

But the last expression is also the value of

$$\int_{-\infty}^{\infty} y(t)[\alpha(t+h+v) - \alpha(t+h)]\,dt.$$

LEMMA 1.5. *If G and H are subsets of $D_{p(t)}$, a necessary and sufficient condition that every element of H should be the limit of a sequence of finite linear combinations of elements of G is that every functional, linear on $D_{p(t)}$, and zero for every element of G, is zero also for every element of H.*

The necessity of the condition is obvious. The sufficiency, with $D_{p(t)}$ replaced by any real normed vectorial space, is a trivial extension of a theorem of H. Hahn.[11] That the condition is sufficient also for $D_{p(t)}$ is a simple consequence of the fact that $D_{p(t)}$ can be regarded as the topological product of two real normed vectorial spaces.[12]

2. Lemmas on Fourier transforms

In this section we collect for reference a number of known facts concerning Fourier transforms and generalized Fourier transforms.

DEFINITION 2.1.[13] *If, for $a \leqq x \leqq b$, $\kappa(x)$ is of class C^n (n a non-negative integer) and $\gamma(x)$ is continuous, we write*

$$d^n f(x) = \kappa(x)\,d^n\gamma(x) \qquad (a \leqq x \leqq b)$$

---

[11] S. Banach, op. cit., p. 58 (Theorem 7).

[12] If E is a real normed vectorial space, with elements $x, y, \cdots$, consider $E^2 = E \times E$, the space of all ordered pairs (x, y) with the operations defined in the usual way (S. Banach, op. cit., p. 182), and $\| (x, y) \| = \| x \| + \| y \|$. E^2 becomes a complex normed vectorial space if we define $(a + ib)(x, y) = (ax - by, bx + ay)$ (a, b real). Then Hahn's theorem is a consequence of a lemma (cf. S. Banach, op. cit., p. 57) which holds in E^2 in the form: *If G is a complex linear subspace of E^2, and the distance between G and (x_0, y_0) of E^2 is $2\delta > 0$, then there exists a linear functional on E^2, zero on G, and not zero at (x_0, y_0).* For, the lower bound, for (x, y) in G, of one of $\| x_0 - x \|$ and $\| y_0 - y \|$ must be not less than δ - say $\| x_0 - x \| \geqq \delta$. If G_R is the real normed vectorial space of all x for (x, y) in G, there is, by the lemma, a real-valued linear functional $f(x)$ on E with $f(x_0) = 1$, $f(x) = 0$ for x in G_R. Define $f((x, y)) = f(x) + if(y)$ for (x, y) in E^2. Then $f((x, y))$ is clearly linear on E^2; and $f((x_0, y_0)) = 1 + if(y_0) \neq 0$; while $-i(x, y)$ is in G if (x, y) is, so that y is in G_R if (x, y) is in G, and so $f(x) = f(y) = f((x, y)) = 0$ for (x, y) in G.

Compare H. F. Bohnenblust and A. Sobczyk, *Extentions of functionals on complex linear spaces*, Bulletin of the American Mathematical Society, 44 (1938), 91–93.

[13] *B*, p. 123.

as an abbreviation for

$$f(x) = \kappa(x)\gamma(x) + \sum_{j=1}^{n} (-1)^j \binom{n}{j} \frac{1}{(j-1)!} \int_0^x (x-t)^{j-1} \kappa^{(j)}(t)\gamma(t)\,dt + P_{n-1}(x) \qquad (a \leqq x \leqq b), \tag{2.1}$$

where $P_{n-1}(x)$ is a polynomial of degree[14] $n-1$. *If* (2.1) *holds and $f(x)$ is itself a polynomial of degree $n-1$, we write*

$$\kappa(x)\, d^n\gamma(x) = 0 \qquad (a \leqq x \leqq b).$$

LEMMA 2.1.[15] *If, for $a \leqq x \leqq b$, $\kappa(x)\, d^n\gamma(x) = 0$, and $\kappa(x)$ is nowhere zero, then $\gamma(x)$ is equal on (a, b) to a polynomial of degree $n-1$.*

LEMMA 2.2.[16] *If the functions $\psi_n(x)$ converge uniformly on (a, b) to $\psi(x)$ $(n \to \infty)$, and if functions $\varphi_n(x)$ are defined so that they satisfy the relations*

$$d^k\varphi_n(x) = \chi(x)\, d^k\psi_n(x) \qquad (a \leqq x \leqq b;\ n = 1, 2, \cdots),$$

then there exist polynomials $P_n(x)$ of degree $k-1$ such that

$$\lim_{n\to\infty} [\varphi_n(x) + P_n(x)] = \varphi(x)$$

exists on (a, b), and satisfies

$$d^k\varphi(x) = \chi(x)\, d^k\psi(x) \qquad (a \leqq x \leqq b).$$

DEFINITION 2.2.[17] *If $|\, g(x)\,|/(1 + |\, x\,|)^k$ (k a non-negative integer) is integrable on $(-\infty, \infty)$, $L_0(x, t) \equiv 0$, and $L_k(x, t)$ denotes the function*

$$L_k(x, t) = \begin{cases} \sum_{n=0}^{k-1} \dfrac{(-ixt)^n}{n!}, & |t| \leqq 1, \\ 0, & |t| > 1, \end{cases} \qquad (k = 1, 2, \cdots),$$

the function

$$\gamma(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(t)\, \frac{e^{-ixt} - L_k(x, t)}{(-it)^k}\, dt$$

is called the k-transform of $g(x)$.

A 0-transform is a Fourier transform in the usual sense.

LEMMA 2.3.[18] *If $g(x)$ and $g_n(x)$ have k-transforms $\gamma(x)$, $\gamma_n(x)$, and if*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\, g(x) - g_n(x)\,|\, (1 + |\, x\,|^k)^{-1}\, dx = 0,$$

---

[14] By "a polynomial of degree k" we shall always mean a polynomial of degree at most k.
[15] *B*, p. 125, 6.
[16] *B*, p. 124, 4.
[17] *B*, p. 112.
[18] *B*, p. 126, 9.

there exist polynomials $P_n(x)$ *of degree* $k-1$ *such that*

$$\gamma(x) = \lim_{n\to\infty}[\gamma_n(x) + P_n(x)],$$

uniformly on every finite interval.

LEMMA 2.4.[19] *If* $g(x)$ *and* $g(x+a)$ *have k-transforms* $\gamma(x)$ *and* $\gamma_a(x)$, *then*

$$d^k\gamma_a(x) = e^{iax}\,d^k\gamma(x).$$

LEMMA 2.5.[20] *If* $K(x)$ *satisfies*

$$|K(x)| < A(1+|x|)^{-k-2} \qquad (-\infty < x < \infty), \tag{2.2}$$

and $\kappa(x)$ *is the Fourier transform of* $K(x)$; *if* $g(x)$ *satisfies the conditions of Definition* 2.2, *and* $\gamma(x)$ *is its k-transform*; *and if*

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} K(x-t)g(t)\,dt;$$

then $F(x)$ *has a k-transform* $\varphi(x)$, *and*

$$d^k\varphi(x) = \kappa(x)\,d^k\gamma(x). \tag{2.3}$$

LEMMA 2.6.[21] *If* $K(x)$ *belongs to* $L(-\infty, \infty)$ *and* $\kappa(x)$ *is its Fourier transform, and if* $\kappa(t)$ *is of class* C^n *on* $(-\infty, \infty)$ *and vanishes outside the finite interval* $(-a, a)$, *then*

$$K(x) = O(|x|^{-n}) \qquad (|x| \to \infty). \tag{2.4}$$

LEMMA 2.7.[22] *If* $s(t)$ *vanishes outside the interval* (α, β) *and has an absolutely convergent Fourier series over an interval* $(-A, A)$ *containing* (α, β); *if* $\Gamma(t)$ *is the Fourier transform of a function belonging to* $L(-\infty, \infty)$, *and* $\Gamma(t)$ *is nowhere zero on* (α, β); *then for a sufficiently large number a, the function*

$$m(t) = \begin{cases} s(t)(1 - a^{-1}|t|)/\Gamma(t), & \alpha \leqq t \leqq \beta, \\ 0, & \text{elsewhere,} \end{cases} \tag{2.5}$$

is the Fourier transform of a function $M(x)$ *of* $L(-\infty, \infty)$.

3. The closure theorem

We shall denote by $C_{p(t)}$ the subset of $D_{p(t)}$ consisting of all functions $x(t)$ of $D_{p(t)}$ which are continuous and have $L_+(x) = L_-(x) = 0$. We say that a subset G of $D_{p(t)}$ is closed in $C_{p(t)}$ if every element of $C_{p(t)}$ is the limit, in the sense of the metric in $D_{p(t)}$, of a sequence of finite linear combinations (with complex constant coefficients) of elements of G.

---

[19] *B*, p. 126, 10.

[20] *B*, p. 127, Satz 40.

[21] *B*, p. 51.

[22] *W*, p. 95. Lemma 2.7 is a collection of facts established by Wiener in the proof of his Lemma 6_{20}.

THEOREM 3.1. *Let $g(t)$ belong to $D_{p(t)}$, with $1/p(t) = O(|t|^n)$ $(|t| \to \infty)$ for an integer $n \geqq 0$, and with $p(t)/p(2t)$ uniformly bounded on $(-\infty, \infty)$. Let $\gamma(t)$ be the $(n+2)$-transform of $g(t)$. A necessary and sufficient condition for the set G of all functions $g(t+u)$ $(-\infty < u < \infty)$ to be closed in $C_{p(t)}$ is that there should exist no interval on which $\gamma(t)$ is equal to a polynomial of degree $n+1$.*

The condition that $p(t)/p(2t)$ is bounded excludes functions of highly irregular decrease, which would cause difficulties.[23] It also implies that $1/p(t) = O(|t|^n)$ for some $n \geqq 0$. The necessity of the condition of the theorem can be disposed of at once. Suppose, in fact, that $d^{n+2}\gamma(t) = 0$ on (β_1, β_2) (in the notation of Definition 2.1). Let $h(t)$ belong to $C_{p(t)}$ and have an $(n+2)$-transform $\theta(t)$ such that $d^{n+2}\theta(t) \not\equiv 0$ on (β_1, β_2). Then $h(t)$ is not the limit of a sequence of linear combinations of elements of G.

To establish this, we assume the contrary; then there exist real numbers u_j and complex numbers $A_{j,m}$ such that

$$\lim_{m\to\infty} ||\, G_m(t) - h(t)\,||_{p(t)} = 0, \tag{3.1}$$

where

$$G_m(t) = \sum_{j=1}^{m} A_{j,m} g(t + u_j).$$

It is clear that (3.1) implies that

$$\lim_{m\to\infty} \int_{-\infty}^{\infty} |\, G_m(t) - h(t)\,|\, (1 + |t|^{n+2})^{-1} = 0;$$

we now apply Lemma 2.3, according to which we can determine polynomials $P_m(t)$ of degree $n+1$ such that

$$\lim_{m\to\infty} \left\{ P_m(t) + \sum_{j=1}^{m} A_{j,m} \gamma_{u_j}(t) \right\} = \theta(t), \tag{3.2}$$

uniformly on (β_1, β_2), where $\gamma_{u_j}(t)$ is the $(n+2)$-transform of $g(t+u_j)$. We now apply Lemma 2.2, taking for $\psi_m(t)$ the expression in the brace in (3.2), and for $\psi(t)$ the function $\theta(t)$. By Lemma 2.4,

$$d^{n+2}\psi_m(t) = \sum_{j=1}^{m} A_{j,m} e^{iu_j t} d^{n+2}\gamma(t),$$

and $d^{n+2}\gamma(t) = 0$ on (β_1, β_2) by assumption. The functions $\varphi_n(t)$ of Lemma 2.2 are then polynomials of degree $n+1$, and may be so chosen that they converge on (β_1, β_2) to a function $\varphi(t)$ satisfying

$$d^{n+2}\varphi(t) = d^{n+2}\theta(t) \qquad (\beta_1 \leqq t \leqq \beta_2).$$

---

[23] The condition is satisfied, for example, by functions t^{-k}, $t^{-k}\log t$, etc. $(k \geqq 0)$, by convex decreasing functions, and by J. Karamata's "fonctions à croissance régulière au sens large."

But $\varphi(t)$, as the limit of a sequence of polynomials of degree $n+1$, is itself a polynomial of degree $n+1$, and we have

$$d^{n+2}\theta(t) = 0 \qquad (\beta_1 \leqq t \leqq \beta_2),$$

which is just the relation which $\theta(t)$ was assumed not to satisfy. Hence (3.1) must be impossible for any choice of the u_j and $A_{j,m}$.

We now establish the sufficiency; we suppose, then, that $d^{n+2}\gamma(t) = 0$ is true on no interval. By Lemma 1.5, we have only to show that if T is any linear functional on $D_{p(t)}$, with $T[g(t-u)] = 0$ $(-\infty < u < \infty)$, then $T[x(t)] = 0$ for every $x(t)$ in $C_{p(t)}$.

We have, by Lemma 1.2, if T is such a functional,

$$T[g(t-u)] = \mathcal{BD}\int_{-\infty}^{\infty} g(t-u)\,d\alpha(t) + \sum_{k=-\infty}^{\infty} \varphi(c_k)[g(-u+c_k) - g(-u+c_k-)] + AL_+(g) + BL_-(g) = 0 \qquad (-\infty < u < \infty). \tag{3.3}$$

We write the same relation with u replaced by $u+\delta$, make an obvious change of variable in the integral, and subtract. We obtain, setting $\alpha_\delta(t) = \alpha(t+\delta) - \alpha(t)$,

$$\mathcal{BD}\int_{-\infty}^{\infty} g(t)\,d\alpha_\delta(t+u) + \sum_{k=-\infty}^{\infty} \varphi(c_k)[g(-u+c_k) - g(-u+c_k-) - g(-u+\delta+c_k) + g(-u+\delta+c_k-)] = 0 \quad (-\infty < u < \infty). \tag{3.4}$$

We integrate (3.4) with respect to u over $(h, h+v)$. By Lemma 1.4, we have

$$\int_h^{h+v} du\,\mathcal{BD}\int_{-\infty}^{\infty} g(t)\,d\alpha_\delta(t+u) = \int_{-\infty}^{\infty} g(t)[\alpha_\delta(t+h+v) - \alpha_\delta(t+h)]\,dt.$$

Relation (1.5),

$$\sum_{k=-\infty}^{\infty} |\varphi(c_k)|/p(c_k) < \infty,$$

together with the uniform boundedness of $p(t)/p(2t)$, implies at once that

$$\sum_{k=-\infty}^{\infty} |\varphi(c_k)|/p(c_k+a)$$

converges uniformly over any finite range of values a. Hence the series in (3.4) is uniformly convergent, and may be integrated term by term over $(h, h+v)$. But $g(-u+c_k)$ differs from $g(-u+c_k-)$ at most on a countable set, and hence the integral of the series is zero. We thus have

$$\int_{-\infty}^{\infty} g(t)f(-t-h)\,dt = 0 \qquad (-\infty < h < \infty), \tag{3.5}$$

where $f(-t) = \alpha_\delta(t + v) - \alpha_\delta(t)$. From (1.4) we obtain

$$\int_{-\infty}^{\infty} p(t)^{-1} |\, d\alpha_\delta(t) \,| < \infty,$$

and consequently, by Lemma 1.3,

$$\int_{-\infty}^{\infty} p(t)^{-1} |\, f(-t) \,|\, dt < \infty. \tag{3.6}$$

The function $f(t)$ depends on δ and v; we shall now show that, for any given δ and v, $f(t)$ is almost everywhere zero.

We suppose the contrary. Then $f(t)$, which belongs to $L(-\infty, \infty)$ by (3.6), has a continuous Fourier transform, $\Gamma(t)$, which is not identically zero, and hence is nowhere zero in some interval (β_1, β_2), which we may assume does not contain the point $t = 0$. Let $s(t)$ be a function of class C^{n+4} on $(-\infty, \infty)$, nowhere zero for $\beta_1 < t < \beta_2$, and vanishing identically outside (β_1, β_2). If the interval $(-a, a)$ contains (β_1, β_2) in its interior, the function

$$\kappa(t) = \begin{cases} s(t)(1 - a^{-1} |\, t \,|), & |\, t \,| < a, \\ 0, & |\, t \,| \geqq a, \end{cases}$$

is of class C^{n+4}. The functions $s(t)$ and $\Gamma(t)$ satisfy the hypotheses of Lemma 2.7; hence the function $m(t)$, defined by (2.5)—

$$m(t) = \begin{cases} s(t)(1 - a^{-1} |\, t \,|)/\Gamma(t), & \beta_1 \leqq t \leqq \beta_2, \\ 0, & \text{elsewhere}— \end{cases}$$

is the Fourier transform of $M(x)$ of $L(-\infty, \infty)$. Let

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} M(t) f(x - t)\, dt. \tag{3.7}$$

Then $K(x)$ belongs to $L(-\infty, \infty)$, and the Fourier transform of $K(x)$ is[24] $m(t)\Gamma(t) = \kappa(t)$.

Lemma 2.6 shows that $K(x)$ and $g(x)$ satisfy the hypotheses of Lemma 2.5, with $k = n + 2$; hence

$$d^{n+2}\varphi(t) = \kappa(t)\, d^{n+2}\gamma(t),$$

where $\varphi(t)$ is the $(n + 2)$-transform of

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K(x - t) g(t)\, dt.$$

---

[24] *B*, p. 46.

But we have, using (3.7) and (3.5),

$$F(x) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} g(t)\,dt \int_{-\infty}^{\infty} M(u)f(x-t-u)\,du$$
$$= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} M(u)\,du \int_{-\infty}^{\infty} g(t)f(x-t-u)\,du$$
$$= 0 \qquad (-\infty < x < \infty);$$

the change in the order of integration is justified by (3.6), the properties of $p(t)$, and Fubini's theorem. Hence $\varphi(t) \equiv 0$, and for all t

$$\kappa(t)\,d^{n+2}\gamma(t) = 0.$$

By Lemma 2.1, this implies that on (β_1, β_2), where $\kappa(t) \neq 0$, $\gamma(t)$ is equal to a polynomial of degree $n+1$. This contradicts our hypothesis on $\gamma(t)$; hence the assumption that $f(t)$ was not almost everywhere zero is untenable.

We therefore have $f(t) = 0$ almost everywhere; and $f(t)$ is a function of bounded variation. Let $\alpha^*(t)$, $\alpha_\delta^*(t)$, $f^*(t)$ denote the functions obtained by normalizing[25] $\alpha(t)$, $\alpha_\delta(t)$, $f(t)$. Then $f^*(t) \equiv 0$; this is true for every v. Since

$$f^*(-t) = \alpha_\delta^*(t+v) - \alpha_\delta^*(t),$$

$\alpha_\delta^*(t)$ must be (for each δ) identically constant:

$$\alpha_\delta^*(t) = \alpha^*(t+\delta) - \alpha^*(t) = c(\delta).$$

If we replace t by $t+\delta$ and subtract, we find

$$\alpha^*(t+2\delta) - 2\alpha^*(t+\delta) + \alpha^*(t) = 0 \qquad (-\infty < t < \infty, -\infty < \delta < \infty).$$

Thus $\alpha^*(t)$ is bounded, and both convex and concave, and therefore a linear function.[26] But this is incompatible with

$$\int_{-\infty}^{\infty} p(t)^{-1}\,|\,d\alpha(t)\,| < \infty$$

unless $\alpha^*(t)$ is a constant.

Now, if $x(t)$ belongs to $C_{p(t)}$, $T[x(t)]$ clearly reduces to

$$\int_{-\infty}^{\infty} x(t)\,d\alpha^*(t),$$

which is zero, since $\alpha^*(t)$ is constant. This is the conclusion which we were to reach in order to establish the closure of the set $\{g(t+u)\}$.

Inspection of the proof of Theorem 3.1 shows that we have also established the apparently more general

---

[25] $\alpha(t)$ is normalized if $\alpha(t) = \frac{1}{2}[\alpha(t+) + \alpha(t-)]$ for every t.

[26] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1934, pp. 91–95.

THEOREM 3.2. *If G is a class of functions of $D_{p(t)}$, with $p(t)$ restricted as in Theorem* 3.1, *the set of all translations of elements of G is closed in $C_{p(t)}$ if and only if there is no interval on which all the $(n+2)$-transforms of elements of G are equal to polynomials of degree $n+1$.*

It does not appear to be possible to establish Theorem 3.1 by establishing first Theorem 3.2 for real spaces (for which Lemma 1.5 is simpler), and then approximating simultaneously to the real and imaginary parts of functions of $C_{p(t)}$.

4. Periodic functions

Let D' denote the space D of Definition 1.1, but with its elements defined over $(0, 2\pi)$ instead of $(-1, 1)$, and continued over $(-\infty, \infty)$ so that they are periodic functions of period 2π; let C' denote the set of continuous functions of D'. It is clear that Lemma 1.5 holds, with appropriate modifications, for the space D'; using it, we establish

THEOREM 4.1. *Let $g(t)$ belong to D' and have Fourier coefficients*

$$\gamma_n = \frac{1}{2\pi}\int_0^{2\pi} g(t)e^{-int}\,dt \qquad (n = 0, \pm 1, \pm 2, \cdots). \tag{4.1}$$

A necessary and sufficient condition for the set G of all $g(t+u)$ $(-\infty < u < \infty)$ to be closed in C' is that $\gamma_n \neq 0$ $(n = 0, \pm 1, \pm 2, \cdots)$.

The necessity of the condition is immediate. If $\gamma_k = 0$, let $f(t) = e^{-ikt}$ Then since the Fourier series of $g(t)$ may be integrated term by term after multiplication by $f(t)$, we have

$$\int_0^{2\pi} g(t)f(t-u)\,dt = \sum_{n=-\infty}^{\infty} \gamma_n \int_0^{2\pi} e^{int}e^{-ik(t-u)}\,dt = 0 \qquad (-\infty < u < \infty).$$

Hence the linear functional

$$T[x(t)] = \int_0^{2\pi} x(t)\,e^{-ikt}\,dt$$

is zero for every $x(t)$ in G; but $T[e^{ikt}] \neq 0$.

To establish the sufficiency of the condition, we consider a linear functional on D', vanishing for every $g(t+u)$, and show that it vanishes on C'. By Lemma 1.1, there exist functions $\alpha(t)$ and $\varphi(t)$ such that

$$\mathcal{RS}\int_0^{2\pi} g(t)\,d\alpha(t+u) + \sum_{k=-\infty}^{\infty} \varphi(c_k)\,[g(c_k-u) - g(c_k-u-)] = 0 \tag{4.2}$$

$$(-\infty < u < \infty),$$

where $\alpha(t)$ has bounded variation on $(0, 2\pi)$, and $\sum_{k=-\infty}^{\infty} |\varphi(c_k)| < \infty$. If we

integrate (4.2) with respect to u on $(\delta, \delta + v)$ we then obtain (using a slight modification of Lemma 1.4)

$$\int_0^{2\pi} g(t) f_\delta(t + v)\, dt = 0 \qquad (-\infty < v < \infty), \tag{4.3}$$

where $f_\delta(t) = \alpha(t + \delta) - \alpha(t)$.
By the familiar "Faltung rule" for Fourier series,[27] (4.3) implies that

$$\gamma_n \varphi_{-n} = 0 \qquad (n = 0, \pm 1, \pm 2, \cdots),$$

where

$$\varphi_n = \frac{1}{2\pi} \int_0^{2\pi} f_\delta(t) e^{-int}\, dt.$$

But no γ_n is zero; therefore every φ_n is zero, and $f_\delta(t)$ is zero almost everywhere. It follows as in the proof of Theorem 3.1 that $\alpha^*(t)$, the normalization of $\alpha(t)$, is constant, and hence that the linear functional which vanishes on G vanishes also on C'.

5. The translations of a sequence

Let $\{x_n\}$ $(n = 0, \pm 1, \pm 2, \cdots)$ be a sequence of complex numbers; by a translation of $\{x_n\}$ we mean a sequence $\{y_n\}$ with $y_n = x_{n+k}$. If a class of sequences is made, in some way, into a metric space, we can ask under what conditions the set of translations of a particular sequence will be closed in the space, or in some subspace. We shall state without proof the answer to this question for the analogues of our spaces $D_{p(t)}$; the reasoning involved would be parallel to that which we have used earlier, although considerably simpler; however, a detailed exposition would involve a number of lemmas which we cannot, as we could in §2, quote from the literature.

DEFINITION 5.1. *By c_p we shall mean the complex normed vectorial space of sequences $x = \{x_n\}$ of complex numbers, with the usual definitions of the operations, where $p = \{p_n\}$ is a positive sequence non-decreasing for $n \leqq 0$ and non-increasing for $n \geqq 0$, $\lim_{n\to\pm\infty} p_n x_n$ exists, and the norm is $\| x \|_p = \max_{-\infty \leqq n \leqq \infty} p_n | x_n |$. c_p^0 shall be the subset of elements x of c_p with $\lim_{n\to\pm\infty} p_n x_n = 0$.*

THEOREM 5.1. *Let p_n/p_{2n} be bounded, and $p_n^{-1} = O(|n|^k)$ $(|n| \to \infty)$ (k a non-negative integer). Let $\{a_n\}$ belong to c_p. The set of sequences $\{a_{n+m}\}$ $(-\infty < m < \infty)$ is closed in c_p^0 if and only if the function $\alpha(t)$ defined by*

$$\alpha(t) = \sum_{n=-\infty}^{\infty}{}' a_n (in)^{-k-2} e^{int}$$

is equal on no interval to a polynomial of degree $k + 1$.

PRINCETON UNIVERSITY.

[27] *W*, p. 45.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

CONVEX REGIONS AND PROJECTIONS IN MINKOWSKI SPACES

BY F. BOHNENBLUST

(Received January 15, 1938)

1. Introduction

By a well known theorem of Hahn-Banach every linear functional of norm M defined over a closed linear subspace of a Banach space[1] can be extended linearly to the entire space without increasing its norm. For operations the corresponding problem takes the following form: Given a linear operation $u(x)$ whose domain of definition is a closed linear subspace of a Banach space B_1 and whose range lies in a Banach space B_2, does there exist an operation $U(x)$ defined over B_1, with range in B_2 and which coincide with $u(x)$ over the subspace? How small can the norm of the extended operation be made?

In the particular case where $u(x) = x$, B_1 = domain of definition of u; the existence of an extension $U(x)$ is equivalent to the existence of a complementary subspace or, as F. J. Murray[2] has shown, to the existence of a projection of B_1 on the subspace. Conversely, if $A(X)$ is such a projection and $u(x)$ any linear operation, the linear operation $U(X) = u(A(X))$ is an extension of $u(x)$ and its norm is $\leqq \| u \| \cdot \| A \|$.

The question of extension is thus equivalent to the discussion of best projections, i.e. of projections of least norm. It is relatively easy to exhibit examples of Banach spaces for which certain closed subspaces have no projections; even more, F. J. Murray[3] has shown that this occurs within the function spaces L_p and the sequence spaces l_p! For such subspaces an extension of an operation is not always possible.

If the Banach space is finite dimensional, i.e. if we are dealing with a Minkowski space, of n dimensions say, the existence of projections is trivial but the discussion of the best possible projections (i.e. those with a minimal norm) is interesting and may help to explain why, in certain infinitely dimensional spaces, projections fail to exist. We give in this paper a complete answer to the question of best possible projections in the case where the dimensionality of the subspace on which we project is by 1 less than the dimensionality of the entire space. The results obtained lead to some interesting theorems on convex regions; they are considered in section 6.

---

[1] For the definition of Banach spaces and the theorem of Hahn Banach cf. e.g. Banach, Opérations linéaires.

[2] F. J. Murray, On complementary manifolds and projections in spaces L_p and l_p, Trans, Amer. Math. Soc., vol. 41 (1937), p. 138.

[3] Ibid.

integrate (4.2) with respect to u on $(\delta, \delta + v)$ we then obtain (using a slight modification of Lemma 1.4)

$$\int_0^{2\pi} g(t) f_\delta(t + v)\, dt = 0 \qquad (-\infty < v < \infty), \tag{4.3}$$

where $f_\delta(t) = \alpha(t + \delta) - \alpha(t)$.
By the familiar "Faltung rule" for Fourier series,[27] (4.3) implies that

$$\gamma_n \varphi_{-n} = 0 \qquad (n = 0, \pm 1, \pm 2, \cdots),$$

where

$$\varphi_n = \frac{1}{2\pi} \int_0^{2\pi} f_\delta(t) e^{-int}\, dt.$$

But no γ_n is zero; therefore every φ_n is zero, and $f_\delta(t)$ is zero almost everywhere. It follows as in the proof of Theorem 3.1 that $\alpha^*(t)$, the normalization of $\alpha(t)$, is constant, and hence that the linear functional which vanishes on G vanishes also on C'.

5. The translations of a sequence

Let $\{x_n\}$ $(n = 0, \pm 1, \pm 2, \cdots)$ be a sequence of complex numbers; by a translation of $\{x_n\}$ we mean a sequence $\{y_n\}$ with $y_n = x_{n+k}$. If a class of sequences is made, in some way, into a metric space, we can ask under what conditions the set of translations of a particular sequence will be closed in the space, or in some subspace. We shall state without proof the answer to this question for the analogues of our spaces $D_{p(t)}$; the reasoning involved would be parallel to that which we have used earlier, although considerably simpler; however, a detailed exposition would involve a number of lemmas which we cannot, as we could in §2, quote from the literature.

DEFINITION 5.1. *By c_p we shall mean the complex normed vectorial space of sequences $x = \{x_n\}$ of complex numbers, with the usual definitions of the operations, where $p = \{p_n\}$ is a positive sequence non-decreasing for $n \leqq 0$ and non-increasing for $n \geqq 0$, $\lim_{n \to \pm\infty} p_n x_n$ exists, and the norm is $\| x \|_p = \max_{-\infty \leqq n \leqq \infty} p_n | x_n |$. c_p^0 shall be the subset of elements x of c_p with $\lim_{n \to \pm\infty} p_n x_n = 0$.*

THEOREM 5.1. *Let p_n/p_{2n} be bounded, and $p_n^{-1} = O(|n|^k)$ $(|n| \to \infty)$ (k a non-negative integer). Let $\{a_n\}$ belong to c_p. The set of sequences $\{a_{n+m}\}$ $(-\infty < m < \infty)$ is closed in c_p^0 if and only if the function $\alpha(t)$ defined by*

$$\alpha(t) = \sum_{n=-\infty}^{\infty}{}' a_n (in)^{-k-2} e^{int}$$

is equal on no interval to a polynomial of degree $k + 1$.

PRINCETON UNIVERSITY.

[27] *W*, p. 45.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

CONVEX REGIONS AND PROJECTIONS IN MINKOWSKI SPACES

BY F. BOHNENBLUST

(Received January 15, 1938)

1. Introduction

By a well known theorem of Hahn-Banach every linear functional of norm M defined over a closed linear subspace of a Banach space[1] can be extended linearly to the entire space without increasing its norm. For operations the corresponding problem takes the following form: Given a linear operation $u(x)$ whose domain of definition is a closed linear subspace of a Banach space B_1 and whose range lies in a Banach space B_2, does there exist an operation $U(x)$ defined over B_1, with range in B_2 and which coincide with $u(x)$ over the subspace? How small can the norm of the extended operation be made?

In the particular case where $u(x) = x$, B_1 = domain of definition of u; the existence of an extension $U(x)$ is equivalent to the existence of a complementary subspace or, as F. J. Murray[2] has shown, to the existence of a projection of B_1 on the subspace. Conversely, if $A(X)$ is such a projection and $u(x)$ any linear operation, the linear operation $U(X) = u(A(X))$ is an extension of $u(x)$ and its norm is $\leqq \| u \| \cdot \| A \|$.

The question of extension is thus equivalent to the discussion of best projections, i.e. of projections of least norm. It is relatively easy to exhibit examples of Banach spaces for which certain closed subspaces have no projections; even more, F. J. Murray[3] has shown that this occurs within the function spaces L_p and the sequence spaces l_p! For such subspaces an extension of an operation is not always possible.

If the Banach space is finite dimensional, i.e. if we are dealing with a Minkowski space, of n dimensions say, the existence of projections is trivial but the discussion of the best possible projections (i.e. those with a minimal norm) is interesting and may help to explain why, in certain infinitely dimensional spaces, projections fail to exist. We give in this paper a complete answer to the question of best possible projections in the case where the dimensionality of the subspace on which we project is by 1 less than the dimensionality of the entire space. The results obtained lead to some interesting theorems on convex regions; they are considered in section 6.

---

[1] For the definition of Banach spaces and the theorem of Hahn Banach cf. e.g. Banach, Opérations linéaires.

[2] F. J. Murray, On complementary manifolds and projections in spaces L_p and l_p, Trans, Amer. Math. Soc., vol. 41 (1937), p. 138.

[3] Ibid.

Let L' be an n-dimensional linear space with a norm $\| X \|$, and L an $(n-1)$-dimensional linear subspace of L'. Elements of L' will be denoted by capital letters $X, Y, \cdots$; those of the subspace by small letters and real numbers by Greek letters. We shall consider exclusively projections of L' on L, they will be denoted by $A, B, \cdots$; their norms by the corresponding Greek letters $\alpha, \beta \cdots$. We find it convenient, furthermore, to denote by $L(X)$ one of the two linear functionals or forms of norm 1, which vanish over L.

THEOREM 1. *For any space L' and any subspace L there exists a projection $A(X)$ of norm $\alpha \leqq 2(n-1)/n$.*

For $n = 1$ this theorem is a trivial statement. It will be proved by induction in the general case; accordingly we shall assume in the sequel that Theorem 1 has been verified for all spaces of dimension $< n$.

2. Special cases.

2.1. An element X of norm 1 will be called maximal with respect to a projection A of norm α if $\| A(X) \| = \alpha$; positive maximal if furthermore $L(X) \geqq 0$.

LEMMA 2.1. *If for a projection $A(X)$, there exists n positive maximal elements $X_1, X_2, \cdots, X_n$ and n non-negative numbers $\rho_1, \rho_2, \cdots, \rho_n$ not all zero, such that $A(\sum \rho_i X_i) = 0$, then the norm α of A is $\leqq 2(n-1)/n$.*

Proof. We may as well assume $\alpha > 1$. Then no maximal element of A can lie in the subspace L, in particular all $L(X_i) > 0$. The elements

$$y_{ij} = \frac{1}{L(X_i) + L(X_j)} [L(X_j)X_i - L(X_i)X_j]$$

belong to the subspace L and to the segments $X_i, -X_j$. Thus we have

$$\| A(y_{ij}) \| = \| y_{ij} \| \leqq 1.$$

On the other hand we can express the $A(X_i)$ in terms of the y_{ij}:

$$A(X_i) \sum_{j=1}^{n} \rho_j L(X_j) = \sum_{j \neq i} \{L(X_i) + L(X_j)\} \rho_j y_{ij}$$

from which follows

$$\alpha \sum_j \rho_j L(X_j) \leqq \sum_{j \neq i} \{L(X_i) + L(X_j)\} \rho_j$$

and therefore by multiplying by $\rho_i / L(X_i)$ and summing with respect to i:

$$(\alpha - 1) \sum_{i,j} \rho_i \rho_j \frac{L(X_j)}{L(X_i)} \leqq \sum_{i,j} \rho_i \rho_j - 2 \sum_i \rho_i^2.$$

But $\dfrac{L(X_j)}{L(X_i)} + \dfrac{L(X_i)}{L(X_j)} \geqq 2$, and thus

$$2 \sum_i \rho_i^2 \leqq (2 - \alpha) \sum \rho_i \rho_j.$$

Taking into account that by Cauchy's inequality $\sum \rho_i \rho_j \leqq n \cdot \sum \rho_i^2$ we obtain finally

$$\alpha \leqq 2(n-1)/n.$$

2.2. Let us consider next the space $l_1^{(n)}$, i.e. the space of elements $Y = \{\eta_i\}$, $i = 1, \cdots, n$ with the norm $\sum |\eta_i|$. Let Y_k be the elements $\eta_i = \delta_{ik}$. Any projection A of $l_1^{(n)}$ on the subspace $\sum \eta_i \xi_i = 0$, where the $\xi_i \geqq 0$, not all zero, is of the form

$$A(Y) = \sum_i (\eta_i - \alpha_i \sum_j \eta_j \xi_j) Y_i$$

where $\sum \alpha_i \xi_i = 1$. The norm α is obviously given by

$$\alpha = \operatorname*{Max}_k \sum_i |\delta_{ik} - \alpha_i \xi_k|$$

and the positive maximal elements of A are some of the Y_k and certain linear combinations of only those Y_k which are maximal. Assume now that A is a projection, whose norm is a minimum for a given set of ξ_i.[4] We distinguish two cases:

(i) all Y_k are maximal
(ii) at least one of the Y_k is not maximal,

which we proceed to discuss separately.

Case (i). In this case no α_i can be negative. Assume to the contrary that $\alpha_1 < 0$. Since $\sum \alpha_i \xi_i = 1$, at least one α_k is positive, say $\alpha_2 > 0$. We have

$$\alpha = |1 - \alpha_1 \xi_1| + \xi_1 \{|\alpha_2| + |\alpha_3| + \cdots + |\alpha_n|\} = 1 + \xi_1 \sum |\alpha_i|$$

and

$$\alpha = |1 - \alpha_2 \xi_2| + \xi_2 \{|\alpha_1| + |\alpha_3| + \cdots + |\alpha_n|\} < 1 + \xi_2 \sum |\alpha_i|$$

and therefore $\xi_2 > \xi_1$. The projection could then be improved by increasing α_1 and by decreasing α_2 so as to maintain the relation $\sum \alpha_j \xi_i = 1$. Thus $\alpha_1 < 0$ leads to a contradiction and we must have $\alpha_i \geqq 0$. But

$$A(\sum \alpha_k Y_k) = 0$$

and therefore it follows by applying lemma 2.1 that the norm of A is $\leqq 2(n-1)/n$.

Case (ii). We prove first

Lemma 2.2. *If A and B of norm α and β respectively are two projections such that*

$$\| B(X) \| < \alpha \text{ for every maximal point } X \text{ of } A,$$

then there exists a projection $C(X)$, whose norm $\gamma < \alpha$.

[4] The existence of minimal projections follows of course from the local compactness of Minkowski spaces.

Proof. Let $\alpha' = \text{Max} \mid\mid B(X) \mid\mid$, when X ranges over the closed set of the maximal points of A. Of course $\alpha' < \alpha$. Let M be the subset of S (= the boundary of the unit sphere) for which $\mid\mid B(X) \mid\mid \leqq (\alpha + \alpha')/2$, and N its complementary set with respect to S. The closure $\bar{N}$ of N includes no maximal point of A; thus, if α'' denotes the $\text{Max} \mid\mid A(X) \mid\mid$ on $\bar{N}$, we have $\alpha'' < \alpha$.

The operation

$$C(X) = (1 - \lambda) \cdot A(X) + \lambda \cdot B(X)$$

is obviously a projection on L. For $0 < \lambda < \text{Min}\,\{1, (\alpha - \alpha'')/\beta\}$ the norm γ of C is $< \alpha$. For if X is a point of M

$$\mid\mid C(X) \mid\mid \leqq (1 - \lambda)\alpha + \lambda \cdot (\alpha + \alpha')/2 < \alpha$$

and if X belongs to N

$$\mid\mid C(X) \mid\mid \leqq \alpha'' + \lambda \cdot \beta < \alpha.$$

Thus the maximal points of a best projection cannot all be improved by any other projection. We return to Case (ii). If not all Y_k are maximal, the space spanned by the maximal points is of dimension smaller than n and thus by induction from the main theorem there exists a projection in this space whose norm is $\leqq 2(n - 2)/(n - 1) < 2(n - 1)/n$. Thus by lemma 2.2 the norm of the best projection $A(X)$ must be $< 2(n - 1)/n$.

In the special case $l_1^{(n)}$ the theorem is verified. We state this result is an analytic form:

THEOREM 2. *To any set of non-negative numbers, ξ_i, $i = 1, \cdots, n$, not all zero, there exists α_i $(i = 1, 2, \cdots, n)$, such that*

1) $\sum \alpha_i \xi_i = 1$

2) $\sum_i \mid \delta_{ik} - \alpha_i \xi_k \mid \leqq 2(n - 1)/n$ *for every* $k = 1, 2, \cdots, n$.

3. A theorem on convex regions.

THEOREM 3. *Let C_α $(k = 1, 2, \cdots)$ be any family of convex, compact regions in an $(n - 1)$-dimensional space such that any n of them have at least one point in common, then the intersection of all C_α is not empty.*

Proofs of this theorem have been given by several authors.[5]

4. General case.

The projections of L' on L form a $(n - 1)$-dimensional space **A**; any A can be represented by

$$A(X) = \sum \xi_i A_i(X), \qquad \sum \xi_i = 1. \tag{4.1}$$

[5] Cf. e.g. Bonessen und Fenchel, Theorie der konvexen Körper, Ergebnisse der Mathematik, Nr. 1, vol. 3, p. 3.

The $A_i(X)$ are projections characterized by $A_i(X_i) = 0$, where the X_i are any n linearly independent elements of the space L' which do not lie in L. Conversely for any ξ_i such that $\sum \xi_i = 1$, $A(X)$ as defined by (4.1) is a projection.

To any point X on the unit sphere which does not lie in the subspace L, there corresponds in the space **A** of the projections a region $C(X)$. This region shall include all projections A, such that the norm of the projection of the element X is $\leqq 2(n-1)/n$. The region $C(-X)$ is obviously equal to $C(X)$. We restrict our considerations to X, for which $L(X) > 0$. The regions $C(X)$ are all compact and any n of them have a non-vanishing intersection. The first statement is obvious, the second follows from theorem 2. For let $X_1, \cdots, X_n$, be any n points, let $\xi_i = L(X_i)$ and determine the α_i according to this theorem. The operation

$$A(X) = X - L(X) \cdot \sum \alpha_i X_i$$

is a projection such that for every $k = 1, 2, \cdots, n$.

$$\| A(X_k) \| \leqq \sum_i | \delta_{ik} - \xi_k \alpha_i | \leqq 2(n-1)/n.$$

By theorem 3, all regions $C(X)$ have at least one projection in common, its norm is clearly $\leqq 2(n-1)/n$ and theorem 1 has been established.

5. Best examples.

The estimates that we have obtained are best possible ones. It is easy to verify that in the n-dimensional space $l_1^{(n)}$, $Y = \{\eta_i\}$, with the norm $\sum | \eta_i |$; the projection

$$A(Y) = Y - \sum_i \eta_i \sum_k Y_k$$

(where as before $Y_k = \{\delta_{ik}\}$) is the minimal projection of $l_i^{(n)}$ on the plane $\sum \eta_i = 0$. For its norm we obtain exactly

$$\alpha = 2(n-1)/n.$$

6. Applications.

6.1 *The radius and center of a region in a Minkowski space.*

Let R be any bounded region in an n-dimensional Minkowski space L of diameter Δ. It is natural to define the radius of such a region as the expression

$$\rho = \operatorname*{Min}_{x_0 \in L} (\text{l.u.b.}_{x \in R} \| x - x_0 \|).$$

Any point x_0 for which the minimum is attained may be called a center of R. The diameter and radius are unchanged if R is increased to the smallest closed convex region containing R. We assume therefore R to be closed and convex. The computation of ρ in the case of a euclidean space was first carried through

by H. W. E. Jung[6] who proved that any closed, convex region in an n-dimensional space is contained in a euclidean sphere of radius $\left[\frac{n}{2(n+1)}\right]^{\frac{1}{2}} \cdot \Delta$. For a general metric the result is given by

THEOREM 6. *The radius ρ of any region R of diameter Δ in an n-dimensional Minkowski space is*

$$\rho \leqq \frac{n}{n+1} \Delta$$

and the factor $(n+1)/n$ is the best possible one.

Consider the $(n+1)$-dimensional space L', which is the direct sum of L and the one-dimensional linear space L_1. Any point X of L' is thus of the form

$$X = (\xi, x)$$

when ξ is a real number, and x an element of L. In L' we consider the norm $|||\, X \,|||$, induced by the convex of $(1, R)$, $(-1, -R)$ as the unit sphere. (We make the inessential assumption that R contains at least one interior point). For any point x of L such that $|||\, (0, x) \,||| = 1$ we have

$$(0, x) = \frac{(+1, x_1) + (-1, -x_2)}{2} = \frac{(0, x_1 - x_2)}{2}$$

where both x_1 and x_2 lie in R. Thus for any $|||\, (0, x) \,||| = 1$ it follows $\| x \| \leqq \Delta/2$ and thus more generally

$$\| x \| \leqq \Delta/2 \, |||\, (0, x) \,|||.$$

By Theorem 1, there exists a projection A of L' on L of norm $\leqq 2n/(n+1)$. Let $(1, x_0)$ be the point in $(1, x)$ whose projection is the origin. Then for any x in R we have

$$\| x - x_0 \| \leqq \tfrac{1}{2}\Delta \, |||(0, x - x_0)||| = \tfrac{1}{2}\Delta \, |||\, A((1, x))|||$$

$$= \Delta \cdot \frac{n}{n+1} \, |||(1, x)||| = \frac{n}{n+1} \cdot \Delta$$

and thus

$$\rho \leqq \Delta \cdot \frac{n}{n+1}.$$

That this factor $n/(n+1)$ is a best result, follows immediately from the examples of section 5. It is interesting to notice that Theorem 6 can be proved directly without making use of our general Theorem 1. We outline briefly such a proof.

---

[6] H. W. E. Jung, Über die kleinste Kugel, die eine räumliche Figur einschliesst, J. für reine und angew. Math., vol. 123 (1901) p. 241.

6.2. *The eccentricity of a convex region and the affine center.*

For any ϵ, $\epsilon \geqq 0$ denote by $S(\epsilon)$ the set containing all elements x such that

$$|\, 2f(x) - (\bar{f} + f)\,| \leqq \epsilon(\bar{f} - f) \tag{6.1}$$

where $f(x)$ is any linear functional and $\bar{f}$ and f its maximum, respectively its minimum as x ranges over R. Obviously $S(\epsilon_1) \subseteq S(\epsilon_2)$, if $\epsilon_1 \leqq \epsilon_2$ and furthermore

$$S(\epsilon) = \prod_{\delta > \epsilon} S(\delta). \tag{6.2}$$

In particular $S(1) = R$. The set of values ϵ, for which $S(\epsilon)$ is not empty is closed, let E be its minimal value. E will be called the eccentricity, and any x in $S(\mathrm{E})$ an affine center, if $\mathrm{E} = 0$, then R is symmetric with respect to a center.

LEMMA 6.1. *The eccentricity of any region R is* $\leqq \dfrac{n-1}{n+1}$.

The proof of this theorem is based on the fact that the center of gravity of the region R lies always in $S\left(\dfrac{n-1}{n+1}\right)$.[7]

Introducing now in the space R a new metric

$$|||\, x \,||| = \operatorname*{Max}_{f \not\equiv 0} \frac{|f(x)|}{(\bar{f} - f)}$$ [8]

we verify next that

$$||\, x \,|| \leqq |||\, x \,|||\, \Delta. \tag{6.3}$$

For $x = 0$ (6.3) is trivial, we may assume therefore that $x \neq 0$. The regions R and $R' = R + x\Delta/||\, x \,||$ have only boundary points in common if any; there exists thus a linear form $f(x) \not\equiv 0$ which separates them, in the sense that $f(x_1) \leqq f(x')$ for any x_1 in R and any x' in R'. In other words $f(x_1) \leqq f(x_2) + f(x)\Delta/||\, x \,||$ for any two elements x_1 and x_2 in R, i.e.

$$\bar{f} \leqq f + f(x)\Delta/||\, x \,||$$

which is equivalent to (6.3). But if x_0 is an affine center of R the inequality

$$|\, 2f(x_0) - (\bar{f} + f)\,| \leqq \mathrm{E}(\bar{f} - f)$$

holds and it implies that

$$|\, 2f(x_0) - 2f(x)\,| \leqq (\mathrm{E} + 1)(\bar{f} - f).$$

For the form which maximize $|f(x - x_0)|$ the last inequality states that $|||\, x - x_0 \,||| \leqq (\mathrm{E} + 1)/2$ and the following theorem is verified:

THEOREM 6.1. *The radius of any region of diameter* Δ *is* $\leqq (\mathrm{E} + 1)\Delta/2$.

Combining this result with lemma 6.1 we obtain theorem 6.

---

[7] Bonessen und Fenchel, Op. cit., p. 52.

[8] Bonessen und Fenchel, Op. cit., p. 53.

7. Banach spaces.

For Banach spaces theorem 6 becomes the trivial remark that the radius of any region is less than or equal to its diameter. We remark though that examples can be constructed for which the equality holds. The theorem 1 on projections is not so trivial. The space L on which the entire space L' is to be projected is taken to be the set of all elements x of the space for which a fixed functional $f \neq 0$ vanishes. For such subspaces there always exist projections of norm $\leqq 2 + \epsilon$, where ϵ is an arbitrary but positive number. Whether such a theorem holds for the limiting case $\epsilon = 0$ could not be established for the most general Banach spaces.

PRINCETON UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

THE PLATEAU PROBLEM FOR MINIMAL SURFACES WHICH ARE RELATIVE MINIMA

BY MAX SHIFFMAN

(Received September 30, 1937)

Introduction

In a previous issue of these Annals,[1] Courant has solved the problem of Plateau for an arbitrary number of boundaries. This solution, as well as Radò's solution for the case of a single boundary and Douglas's solution for one and two boundaries,[2] solves at the same time the problem of least area and yields the absolute minimum. To insure the solvability of the problem by a non-degenerate surface, assumptions in the form of inequalities must be made. These inequalities express the condition that the lower limit of the area given by non-degenerate surfaces be less than that given by degenerate surfaces. It turns out in the paper quoted above (and also in Douglas's papers) that the inequalities also serve as important tools in the course of the solution.

The purpose of the present paper, as has already been announced in Courant's paper,[3] is to slightly modify the method of Courant in such a way that not only absolute minima but also certain types of relative minima fall within its scope. This goal is attained by replacing the inequalities described above by a more general inequality which likewise excludes degenerations. This new inequality is based on the requirement that only those surfaces are admitted in the variational problem for which a certain quantity called its diameter is not smaller than a fixed positive bound.

The most obvious example in which it is desirable to have an existence proof for relative minima is that of a catenoid where the two end points are nearly conjugate.[4] To prove, in particular cases, the validity of our generalized inequality, one can use the method of continuity developed by Courant for similar purposes.[5] In this way, the existence of minimal surfaces which are relative minima can be established for numerous examples of contours.

In order to give the essentials of the method, we deal first with two bounda-

---

[1] Courant, *"Plateau's Problem* and *Dirichlet's Principle"*, Annals of Math., vol. 38, no. 3, July, 1937, pp. 679–724.

[2] See Douglas's paper, *The Problem of Plateau*, Bull. Am. Math. Soc. (1933), pp. 227–251, and the excellent report of Rado, *The Problem of Plateau*, Ergebnisse der Math., vol. II, no. 2 (1933) for complete references.

[3] L. c., p. 684 and footnote 24, p. 702.

[4] See Bliss, *The Calculus of Variations*, Chap. IV, for a beautiful presentation of the catenoid problem.

[5] L. c., p. 695.

ries, the case of several boundaries being summarily discussed in (4). To avoid duplications, detailed references to Courant's paper will be made.

1. The Diameter of a Surface. Statement of the Theorem

Let $\mathfrak{x} = \mathfrak{x}(u, v)$ be a surface (in vector form) in $\mathfrak{x}$-space, the parameters u, v ranging over a circular region G of the (u, v)-plane, i.e., over a region bounded by circles in the interior of one larger circle (the unit circle). We define the *diameter* of $\mathfrak{x}(u, v)$ as the greatest lower bound of the diameters[6] of all curves which cannot be reduced to a point on the surface, i.e., of all those curves which are images on $\mathfrak{x}(u, v)$ of the curves in G completely surrounding at least one of the inner circles of G. The diameter is a measure of how far from degeneracy the surface is.

Clearly, if the surfaces $\mathfrak{x}_n(u, v)$ converge uniformly to the surface $\mathfrak{x}(u, v)$, the diameters t_n of $\mathfrak{x}_n$ tend to the diameter t of $\mathfrak{x}$.

Let Γ_1, Γ_2 be two prescribed contours. Consider surfaces $\mathfrak{x} = \mathfrak{x}(u, v)$, the parameters u, v ranging over annular rings with the unit circle as outer boundary, which map the two concentric circular boundaries on the curves Γ_1, Γ_2 respectively, and which are continuous and have piecewise continuous first derivatives.[7] Courant's method begins with the problem of finding, among all such surfaces, one which minimizes the Dirichlet functional

$$\mathfrak{D}(\mathfrak{x}) = \frac{1}{2}\int\int (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\, dudv = \frac{1}{2}\int\int \left(\mathfrak{x}_r^2 + \frac{1}{r^2}\mathfrak{x}_\theta^2\right) rdrd\theta.$$

This problem will here be modified by admitting only surfaces whose diameters are not less than a fixed positive constant.

THEOREM. *Let Γ_1, Γ_2 be two non-intersecting closed Jordan curves, and let α be any positive number. Define d_α as the greatest lower bound of $\mathfrak{D}(\mathfrak{x})$ for surfaces[7] $\mathfrak{x}(u, v)$ bounded by Γ_1, Γ_2 whose diameters are not less than α; and δ_α as the greatest lower bound of $\mathfrak{D}(\mathfrak{x})$ for those surfaces whose diameters are exactly equal to α. Then there exists a minimal surface bounded by Γ_1, Γ_2 if the following inequality is satisfied:*[8] $d_2 < \delta_\alpha$.

Note, from the definition itself, that $d_\alpha \leqq \delta_\alpha$. The theorem requires the inequality sign to hold.

2. Lemmas on the Dirichlet Functional

LEMMA 1.[9] *Let s be the smallest distance between the points of the curve Γ_1 and those of Γ_2. Let $\mathfrak{x}(u, v)$ be a surface of diameter t bounded by Γ_1, Γ_2 and defined*

---

[6] The diameter of a curve is the maximum distance between pairs of its points.

[7] In all that follows, the surfaces considered shall be continuous and have piecewise continuous first derivatives.

[8] The inequality excludes the possibility $d_\alpha = \infty$. It should be noted that α must be less than the diameters of each of the curves Γ_1, Γ_2.

[9] On account of the positive α in our theorem, this lemma will exclude degeneracy in *one* step. It will be found that Courant's considerations on degeneracy and equicontinuity are somewhat simpler for our more general theorem.

over the annular ring G of the (u, v)-plane of inner radius ρ, outer radius ρ'; Then

$$\mathfrak{D}(\mathfrak{x}) \geqq \frac{\pi s^2}{\log \frac{\rho'}{\rho}} + \frac{1}{\pi} t^2 \log \frac{\rho'}{\rho}.$$

PROOF. (a) Integrating along a radial line θ = constant from the inner to the outer circle of G,

$$s \leqq \left| \int_\rho^{\rho'} \mathfrak{x}_r \, dr \right| \leqq \int_\rho^{\rho'} r^{\frac{1}{2}} \mid \mathfrak{x}_r \mid \cdot \, r^{-\frac{1}{2}} dr$$

and

$$s^2 \leqq \int_\rho^{\rho'} \frac{dr}{r} \cdot \int_\rho^{\rho'} r \mathfrak{x}_r^2 \, dr = \log \frac{\rho'}{\rho} \cdot \int_\rho^{\rho'} r \mathfrak{x}_r^2 \, dr.$$

Integrating with respect to θ from 0 to 2π,

$$\int_\rho^{\rho'} \int_0^{2\pi} r \mathfrak{x}_r^2 \, dr d\theta \geqq \frac{2\pi s^2}{\log \frac{\rho'}{\rho}}.$$

(b) Since the diameter of $\mathfrak{x}(u, v)$ is equal to t, on any circle of radius r about the origin there exist two points A, B whose images are at a distance not less than t, i.e., $\mid \mathfrak{x}(A) - \mathfrak{x}(B) \mid \geqq t$. Integrating along this circle between A and B, we have

$$t \leqq \left| \int_A^B \mathfrak{x}_\theta \, d\theta \right| \leqq \int_A^B \mid \mathfrak{x}_\theta \mid d\theta.$$

As there are two arcs from A to B,

$$2t \leqq \int_0^{2\pi} \mid \mathfrak{x}_\theta \mid d\theta$$

and

$$4t^2 \leqq 2\pi \int_0^{2\pi} \mathfrak{x}_\theta^2 \, d\theta.$$

Multiply by $1/r$ and integrate with respect to r from ρ to ρ':

$$\int_\rho^{\rho'} \int_0^{2\pi} \frac{1}{r} \mathfrak{x}_\theta^2 \, dr d\theta \geqq \frac{2}{\pi} t^2 \log \frac{\rho'}{\rho}.$$

The parts (a) and (b), together with the expression for $\mathfrak{D}(\mathfrak{x})$ in polar coordinates, complete the proof of the lemma.

LEMMA 2. *Suppose that the surface $\mathfrak{x}(u, v)$ is defined over an annular ring*

$\rho \leqq r \leqq \rho'$ of the (u, v)-plane, and let t_θ be the oscillation of $\mathfrak{x}(u, v)$ on the radial straight line $\theta =$ constant. Then, for any positive η,

$$t_\theta \leqq \left(\log \frac{\rho'}{\rho} \cdot \frac{2\mathfrak{D}(\mathfrak{x})}{\eta}\right)^{\frac{1}{2}}$$

except for a set of intervals of θ of total length less than η.

PROOF. By virtue of $\frac{1}{2}\int_0^{2\pi}\int_\rho^{\rho'} \mathfrak{x}_r^2 r dr d\theta \leqq \mathfrak{D}(\mathfrak{x})$, it follows that $\int_\rho^{\rho'} \mathfrak{x}_r^2 r dr d\theta \leqq \frac{2\mathfrak{D}(\mathfrak{x})}{\eta}$ except for a set of intervals of θ of total length less than η. From $t_\theta \leqq \int_\rho^{\rho'} |\mathfrak{x}_r| dr$ we have, except for these θ's,

$$t_\theta^2 \leqq \log \frac{\rho'}{\rho} \cdot \int_\rho^{\rho'} \mathfrak{x}_r^2 r dr \leqq \log \frac{\rho'}{\rho} \cdot \frac{2\mathfrak{D}(\mathfrak{x})}{\eta}.$$

The lemma 5, pp. 688, 689, of Courant's paper, the proof of which is a trivial modification of part (b) of the proof of our lemma 1, is also useful. Since its uses are contained in other lemmas of Courant to which we shall refer, it shall not be mentioned explicitly.

3. Proof of the Theorem. Minimizing Sequences

We shall solve the variational problem

$$\mathfrak{D}(\mathfrak{x}) = \text{minimum} = d_\alpha$$

among all surfaces bounded by Γ_1, Γ_2 whose diameters are not less than α.

(a) Let $\mathfrak{x}_1(u, v)$, $\mathfrak{x}_2(u, v)$, $\cdots$ be a minimizing sequence, $\mathfrak{D}(\mathfrak{x}_n) \to d_\alpha$, and let the corresponding annular rings have outer radius 1 and inner radii ρ_1, ρ_2, $\cdots$ respectively. By lemma 1,

$$\mathfrak{D}(\mathfrak{x}_n) \geqq \frac{\pi s^2}{\log \frac{1}{\rho_n}} + \frac{1}{\pi}\alpha^2 \log \frac{1}{\rho_n}.$$

It follows that ρ_n cannot tend to 0 or 1, and a subsequence which we rename $\mathfrak{x}_1$, $\mathfrak{x}_2$, $\cdots$ can be found for which $\rho_n \to \rho \neq 0, 1$.

(b) The boundary values of $\mathfrak{x}_n$ are equicontinuous. For if not, there is a subsequence which we call $\mathfrak{x}_1$, $\mathfrak{x}_2$, $\cdots$, a positive constant γ, and sequences of boundary points P_n, Q_n tending to a point R such that $|\mathfrak{x}_n(P_n) - \mathfrak{x}_n(Q_n)| > \gamma$. As in Courant's paper (lemma 6 and figures 1A, 1B, p. 690, 691), there is a circular arc $\widehat{PQ}$ (depending on n) on which the oscillation of $\mathfrak{x}_n$ is less than[10]

$$\left(\frac{2\pi\,\mathfrak{D}(\mathfrak{x}_n)}{|\log \delta_n|}\right)^{\frac{1}{2}}$$

[10] We choose $\delta_n \to 0$ instead of a fixed δ.

which tends to zero as $n \to \infty$. In particular, the distance between $\mathfrak{x}_n(P)$ and $\mathfrak{x}_n(Q)$ tends to zero, and this implies that the diameter of the image of the arc $\overset{\frown}{QbP}$ complimentary to $\overset{\frown}{PRQ}$ tends to zero. The diameter of the image on $\mathfrak{x}_n$ of the closed curve $\overset{\frown}{PQ}\overset{\frown}{QbP}$ therefore tends to zero, in contradiction to the hypothesis that the diameters of $\mathfrak{x}_n$ are not below α. So the boundary values of $\mathfrak{x}_n$ are equicontinuous.

(c) Since $\mathfrak{D}(\mathfrak{x}_n) \to d_\alpha$ and $\delta_\alpha > d_\alpha$, only a finite number of surfaces of the minimizing sequence can have diameters equal to α, and these may be deleted. We will now show, and this is the core of our existence proof, that it is possible to replace each surface of the minimizing sequence by a potential surface without violating the conditions of our variational problem.

Let $\mathfrak{x}(u, v)$ be a fixed surface of the minimizing sequence and ρ the inner radius of its annular ring. For any value of R between ρ and 1, define $\mathfrak{x}_R(u, v)$ as the surface identical with $\mathfrak{x}(u, v)$ in the annular ring $\rho \leqq r \leqq R$, but equal in the ring $R \leqq r \leqq 1$ to the potential surface having the same boundary values as $\mathfrak{x}(u, v)$ on the circles $r = R$, $r = 1$. Clearly, $\mathfrak{x}_R(u, v)$ is continuous, has piecewise continuous first derivatives, is bounded by Γ_1, Γ_2, and $\mathfrak{D}(\mathfrak{x}_R) \leqq \mathfrak{D}(\mathfrak{x})$. In addition, $\mathfrak{x}_1(u, v) \equiv \mathfrak{x}(u, v)$ while $\mathfrak{x}_\rho(u, v)$ is a potential surface.

We shall show that the diameter t_R of $\mathfrak{x}_R(u, v)$ is a continuous function of R. It suffices to prove that $\mathfrak{x}_{R_n}(u, v)$ converges uniformly to $\mathfrak{x}_R(u, v)$ if $R_n \to R$. If $R \neq 1$, this is a consequence of a well-known theorem of potential theory. Suppose that $R_n \to 1$; then the potential surface $\mathfrak{x}_{R_n}(u, v)$ in the ring $R_n \leqq r \leqq 1$ converges uniformly to the curve represented by $\mathfrak{x}(u, v)$ on $r = 1$, as the following argument proves:[11]

Since $\mathfrak{x}_{R_n}(u, v) = \mathfrak{x}(u, v)$ on the circles $r = R_n$ and $r = 1$, and $\mathfrak{x}(u, v)$ is uniformly continuous, for any ϵ there is an η independent of n such that

$$| \mathfrak{x}_{R_n}(P) - \mathfrak{x}_{R_n}(Q) | < \epsilon/3$$

where P, Q are any two points on $r = R_n$ or $r = 1$ whose distance is less than η. By lemma 2, the diameter of the image on $\mathfrak{x}_{R_n}$ of the radial straight line $\theta =$ constant is less than

$$\left(\log \frac{1}{R_n} \cdot \frac{2\mathfrak{D}(\mathfrak{x})}{\eta}\right)^{\frac{1}{2}},$$

or less than $\epsilon/3$ for sufficiently large n, except for a set of intervals of θ (depending on n) of total length less than η. This permits the partitioning of the ring $R_n \leqq r \leqq 1$ into domains H_1, H_2, $\cdots$, H_l by radial straight lines (depending on n)

$$\theta_1, \theta_2, \cdots, \theta_l, \theta_{l+1} = \theta_1 + 2\pi$$

[11] The idea of this argument was kindly suggested to me by Professor Courant.

on each of which the oscillation of $\mathfrak{x}_{R_n}$ is less than $\epsilon/3$ and for which $\theta_{i+1} - \theta_i < \eta$. The total oscillation of $\mathfrak{x}_{R_n}$ on the boundary of each one of the domains $H_1, H_2, \cdots, H_l$ is less than ϵ. Since $\mathfrak{x}_{R_n}(u, v)$ is a potential surface in these domains, the oscillation of $\mathfrak{x}_{R_n}$ in each of them is less than ϵ. In particular

$$| \mathfrak{x}_{R_n}(r, \theta) - \mathfrak{x}(1, \theta) | < \epsilon$$

for all points (r, θ) in $R_n \leqq r \leqq 1$, provided n is large enough.

It follows that $\mathfrak{x}_{R_n}(u, v)$ converges uniformly to $\mathfrak{x}(u, v)$. The proof that t_R is a continuous function of R in $\rho \leqq R \leqq 1$ is thereby completed.

Since t_1, the diameter of $\mathfrak{x}(u, v)$, is larger than α (the surfaces of the minimizing sequence with diameter equal to α have been deleted), one of the two alternatives occurs: either $t_\rho > \alpha$ or there is an R' for which $t_{R'} = \alpha$. Replace $\mathfrak{x}(u, v)$ by $\mathfrak{x}'(u, v)$ where $\mathfrak{x}'$ is either $\mathfrak{x}_\rho$ or $\mathfrak{x}_{R'}$ according as the first or second alternative occurs. The new surface $\mathfrak{x}'(u, v)$ is an admissible surface for which $\mathfrak{D}(\mathfrak{x}') \leqq \mathfrak{D}(\mathfrak{x})$ and it is either a potential surface or has a diameter equal to α.

Replacing each one of $\mathfrak{x}_1, \mathfrak{x}_2, \cdots$ by $\mathfrak{x}_1', \mathfrak{x}_2', \cdots$ in this way, we again have a minimizing sequence. Only a finite number of these can have diameters equal to α. Deleting them, there remains a minimizing sequence consisting of potential surfaces.

(d) By part (b), the boundary values of $\mathfrak{x}_n'$ are equicontinuous so that a subsequence can be chosen which converges uniformly to a potential surface $\psi(u, v)$ defined over the annular ring $\rho \leqq r \leqq 1$ of the (u, v)-plane. $\psi(u, v)$ is the solution to our variational problem. As in Courant's paper, varying $\psi(u, v)$ will yield the result that $\psi(u, v)$ is a minimal surface[12] q.e.d.

4. The Theorem for Several Boundaries

We now turn to the theorem for the case of an arbitrary number k of non-intersecting closed Jordan curves $\Gamma_1, \cdots, \Gamma_k$. Consider surfaces $\mathfrak{x}(u, v)$ defined over regions G of the (u, v)-plane consisting of $k - 2$ circular holes cut out of the interior of an annular ring (with unit circle as outer boundary), and mapping the k boundaries of G on $\Gamma_1, \cdots, \Gamma_k$ respectively. Defining d_α, δ_α as previously, we have the:

THEOREM. *If $d_\alpha < \delta_\alpha$ there exists a minimal surface bounded by $\Gamma_1, \cdots, \Gamma_k$.*

PROOF. Suppose that $\mathfrak{x}_1, \mathfrak{x}_2, \cdots$ with domains $G_1, G_2 \cdots$ is a minimizing sequence of surfaces of diameters larger than α, $\mathfrak{D}(\mathfrak{x}_n) \to d_\alpha$. If we exclude all possibilities of degeneration, the method of (3) (b), (c), (d) will yield the required result.

Firstly, no circle or group of circles of G_n with radii tending to zero can tend to a point O away from all other circles. For then annular rings contained

[12] The variation of the boundary values, $z(r, \theta) = \psi(r, \varphi)$ where $\varphi = \theta + \epsilon\lambda(r, \theta)$, which Courant performs does not alter the diameter of ψ. The variation of the radius of the inner circle changes it but slightly. For this latter variation, one can also use $z(r, \theta) = \psi(R, \theta)$ where $r = R + \epsilon\lambda(R, \theta)$ and $\lambda(\rho, \theta) = 1, \lambda(1, \theta) = 0$; this does not alter the diameter.

completely in G_n with fixed outer radius ϵ and inner radius ϵ_n tending to zero could be drawn about O, and lemma 1 would yield a contradiction. Secondly, by lemma (c) and figure 3, pp. 701, 702, of Courant's paper, no pair of circles with radii above a positive bound can tend to each other. The remaining possibility of degeneration is that one or more circles of radii tending to zero approach a circle whose radius is above a positive bound. Just as on p. 710, figure 5 of Courant's paper, there is an arc $\widehat{PQ}$ (depending on n) surrounding the small circles on which the oscillation of $\mathfrak{x}_n(u, v)$ tends to zero. According as R is a point of equicontinuity or non-equicontinuity, the diameter of the image on $\mathfrak{x}_n$ of the closed curve $\widehat{PQQbP}$ or $\widehat{PQQb'P}$ respectively tends to zero. This is contrary to hypothesis, and degeneracy cannot occur.

A subsequence can be chosen the domains of which tend to a domain G consisting of an annular ring with $k - 2$ circular holes cut out. The reasoning of (3) (b), (c), (d) establishes the existence of a minimal surface bounded by $\Gamma_1, \cdots, \Gamma_k$.

5. Concluding Remarks

We have made essential use of the condition that α be positive. The theorem also holds if $\alpha = 0$, in which case the inequality $d_0 < \delta_0$ is easily seen to be identical to the inequalities used by Courant. The proof of the theorem for $\alpha = 0$ would proceed as follows: (a) one can establish the continuity of δ_α and d_α as functions of α, especially at $\alpha = 0$;[13] (b) the inequality $d_0 < \delta_0$ then implies $d_\alpha < \delta_\alpha$ for sufficiently small positive α.

Finally, the lower semi-continuity, approximation, and continuity theorems are easily proved for our general case. The approximation theorem is particularly simple for positive α, the requirement being

$$d_\alpha^{(n)} < \delta_\alpha^{(n)}$$

in place of[14]

$$d^{(n)} < d_I^{(n)} + d_{II}^{(n)} - \gamma.$$

NEW YORK UNIVERSITY.

[13] By exactly the method of lower semi-continuity in Courant's paper.

[14] Cf. Courant, l.c., pp. 706, 707.

FIELDS OF PARALLEL VECTORS IN RIEMANNIAN SPACE

By Luther Pfahler Eisenhart

(Received December 17, 1937)

In a V_n of coordinates $x^i (i = 1, \cdots, n)$ and with the fundamental tensor g_{ij}, not necessarily positive definite, the vector-field of components λ^i is said to be a field of parallel vectors, if λ^i satisfy the system of equations

$$\lambda^i_{,j} \equiv \frac{\partial \lambda^i}{\partial x^j} + \lambda^k \left\{ \begin{matrix} i \\ kj \end{matrix} \right\} = 0, \tag{1}$$

where $\left\{ \begin{matrix} i \\ kj \end{matrix} \right\}$ are Christoffel symbols of the second kind formed with respect to the tensor g_{ij}; and $\lambda^i_{,j}$ denotes a covariant derivative of λ^i; the summation convention will be used throughout this paper. In a previous paper[1] we considered this problem, and showed that the determination of whether a given space admits one or more independent fields of parallel vectors is an algebraic problem. In the present paper we determine canonical forms of the fundamental quadratic form of a V_n, admitting independent fields of parallel vectors.

Suppose that λ^i_a are the components of r such fields, where $a(= 1, \cdots, r)$ indicates the field and $i(= 1, \cdots, n)$ the component. If we put

$$c_{ab} = g_{ij} \lambda^i_a \lambda^j_b \qquad (a, b = 1, \cdots, r), \tag{2}$$

the c's are scalars. Differentiating (2) with respect to x^k, and, noting that the covariant derivatives of the g_{ij} are zero, we have in consequence of (1) that the c's are constants. When the fundamental form is indefinite, there is a possibility that some of the parallel fields are null vector fields, that is some, or all, of the constants c_{aa} may be equal to zero. From the form of equations (1) it is seen that any linear combination of the λ's with real constant coefficients are the components of a parallel field. By means of such linear transformations it is possible to transform the matrix $\| c_{ab} \|$, when its rank is $p(\leqq r)$, into one for which we have

$$\begin{gathered} c_{\alpha\alpha} = e_\alpha, \qquad c_{\mu\mu} = 0, \qquad c_{ab} = 0 \\ (\alpha = 1, \cdots, p; \mu = p + 1, \cdots, r; a, b = 1, \cdots, r; a \neq b), \end{gathered} \tag{3}$$

where the e's are $+1$ or -1 as the case may be. Thus in the new vector fields λ^i_α are unit vectors, λ^i_μ are null vectors, and any two fields are mutually orthogonal.

---

[1] *Fields of parallel vectors in a Riemannian geometry.* Transactions of the Amer. Math. Soc., vol. 27 (1925), pp. 563–573; see also the author's *Riemannian Geometry*, pp. 67–72, to which hereafter a reference is given as R. G.

We have shown[2] that a coordinate system exists with respect to which the components of each of these fields are equal to zero except the component with the same subscript and superscript. Accordingly we have, in consequence of (3),

$$(4)\quad g_{\alpha\beta} = g_{\alpha\mu} = g_{\mu\nu} = 0 \qquad (\alpha, \beta = 1, \cdots, p; \alpha \neq \beta; \mu, \nu = p+1, \cdots, r).$$

Since the vectors λ_α^i are unit vectors, we have

$$(5)\quad \lambda_\alpha^\alpha = (e_\alpha g_{\alpha\alpha})^{-\frac{1}{2}}, \qquad \lambda_\alpha^h = 0 \qquad (h = 1, \cdots, n; h \neq \alpha),$$

and

$$(6)\quad \lambda_\mu^\mu = \varphi_\mu, \qquad \lambda_\mu^s = 0 \qquad (s = 1, \cdots, n; s \neq \mu),$$

where the functions φ_μ are thus defined.

We consider first the case of the unit vectors. When the expressions (5) are substituted in (1), we obtain

$$\frac{\partial}{\partial x^j} \log (e_\alpha g_{\alpha\alpha})^{\frac{1}{2}} - \begin{Bmatrix} \alpha \\ j\alpha \end{Bmatrix} = 0 \qquad (\alpha \text{ not summed}),$$

$$\begin{Bmatrix} h \\ j\alpha \end{Bmatrix} = 0 \quad (j, h = 1, \cdots, n; h \neq \alpha).$$

If we multiply the first of these equations by $g_{\alpha k}$ and subtract from it the second multiplied by g_{hk} and summed for h, we get the equivalent set of equations

$$(7)\quad g_{\alpha k} \frac{\partial}{\partial x^j} \log g_{\alpha\alpha} - \frac{\partial g_{jk}}{\partial x^\alpha} - \frac{\partial g_{\alpha k}}{\partial x^j} + \frac{\partial g_{\alpha j}}{\partial x^k} = 0 \qquad \begin{pmatrix} \alpha = 1, \cdots, p; \\ j, k = 1, \cdots, n \end{pmatrix}.$$

When in (7) we put $k = \alpha$, the equation is satisfied identically. When we put $k = \beta = 1, \cdots, p$ $(\beta \neq \alpha)$, we obtain in consequence of (4)

$$(8)\quad \frac{\partial g_{j\beta}}{\partial x^\alpha} - \frac{\partial g_{j\alpha}}{\partial x^\beta} = 0.$$

In particular, when $j = \alpha$, we have

$$(9)\quad \frac{\partial g_{\alpha\alpha}}{\partial x^\beta} = 0 \qquad (\alpha, \beta = 1, \cdots, p; \alpha \neq \beta).$$

If we put $j = \beta$ in (7), we have in consequence of (9) and (4)

$$\frac{\partial g_{\beta k}}{\partial x^\alpha} + \frac{\partial g_{\alpha k}}{\partial x^\beta} = 0.$$

From this equation and (8) it follows that

$$(10)\quad \frac{\partial g_{\alpha\sigma}}{\partial x^\beta} = 0 \qquad (\alpha, \beta = 1, \cdots, p; \sigma = p+1, \cdots, n).$$

[2] R. G., p. 70.

When now we put $j = \alpha$ in (7), the resulting equation may be written in the form

$$\frac{\partial}{\partial x^\alpha}(g_{\alpha\sigma}(e_\alpha g_{\alpha\alpha})^{-\frac{1}{2}}) = e_\alpha \frac{\partial}{\partial x^\sigma}(e_\alpha g_{\alpha\alpha})^{\frac{1}{2}} \quad (\sigma = p+1, \cdots, n). \tag{11}$$

If we define functions

$$(e_\alpha g_{\alpha\alpha})^{\frac{1}{2}} = \frac{\partial \psi_\alpha}{\partial x^\alpha}, \tag{12}$$

in consequence of (9) there is no loss of generality in taking ψ_α as a function of $x^\alpha, x^{p+1}, \cdots, x^n$ at most. Substituting from (12) in (11) and integrating the resulting equation, we have

$$g_{\alpha\sigma} = e_\alpha \frac{\partial \psi_\alpha}{\partial x^\alpha}\left(\frac{\partial \psi_\alpha}{\partial x^\sigma} + \varphi_{\alpha\sigma}\right), \tag{13}$$

where $\varphi_{\alpha\sigma}$ is independent of x^α. In consequence of (10) $\varphi_{\alpha\sigma}$ is a function of $x^{p+1}, \cdots, x^n$ at most.

If in (7) we put $j = \sigma, k = \tau$ for $\sigma, \tau = p+1, \cdots, n$ and substitute from (12) and (13), we obtain

$$e_\alpha \frac{\partial g_{\sigma\tau}}{\partial x^\alpha} = \left(\frac{\partial \psi_\alpha}{\partial x^\sigma} + \varphi_{\alpha\sigma}\right)\frac{\partial^2 \psi_\alpha}{\partial x^\alpha \partial x^\tau} + \left(\frac{\partial \psi_\alpha}{\partial x^\tau} + \varphi_{\alpha\tau}\right)\frac{\partial^2 \psi_\alpha}{\partial x^\alpha \partial x^\sigma} + \frac{\partial \psi_\alpha}{\partial x^\alpha}\left(\frac{\partial \varphi_{\alpha\sigma}}{\partial x^\tau} - \frac{\partial \varphi_{\alpha\tau}}{\partial x^\sigma}\right). \tag{14}$$

Since $g_{\sigma\tau}$ is symmetric in σ and τ, we must have

$$\frac{\partial \varphi_{\alpha\sigma}}{\partial x^\tau} = \frac{\partial \varphi_{\alpha\tau}}{\partial x^\sigma},$$

from which it follows that

$$\varphi_{\alpha\sigma} = \frac{\partial \varphi_\alpha}{\partial x^\sigma},$$

where φ_α thus defined is a function of $x^{p+1}, \cdots, x^n$ at most. Hence φ_α may be incorporated in ψ_α in (12) and (13), and then in all generality we may take $\varphi_{\alpha\sigma} = 0$. Consequently equations (13) become

$$g_{\alpha\sigma} = e_\alpha \frac{\partial \psi_\alpha}{\partial x^\alpha}\frac{\partial \psi_\alpha}{\partial x^\sigma}. \tag{15}$$

Then from (14), since ψ_α contains $x^\alpha, x^{p+1}, \cdots, x^n$ for $\alpha = 1, \cdots, p$, we have

$$g_{\sigma\tau} = \sum_\alpha e_\alpha \frac{\partial \psi_\alpha}{\partial x^\sigma}\frac{\partial \psi_\alpha}{\partial x^\tau} + \varphi_{\sigma\tau}, \tag{16}$$

where the functions $\varphi_{\sigma\tau}$ are independent of $x^1, \cdots, x^p$.

From (12), (15) and (16) it follows that the fundamental quadratic form is

$$\sum_\alpha e_\alpha (d\psi_\alpha)^2 + \varphi_{\sigma\tau}\, dx^\sigma\, dx^\tau.$$

A non-singular transformation of coordinates is defined by

$$x'^{\alpha} = \psi_{\alpha}, \qquad x'^{\sigma} = x^{\sigma} \qquad (\alpha = 1, \cdots, p; \sigma = p+1, \cdots, n),$$

in terms of which the above form becomes

$$\sum_{\alpha} e_{\alpha}(dx^{\alpha})^2 + g_{\sigma\tau}\, dx^{\sigma}\, dx^{\tau}, \tag{17}$$

where $g_{\sigma\tau}$ are functions of $x^{p+1}, \cdots, x^{n}$, at most. In the new coordinate system

$$\lambda^{i}_{\alpha} = \delta^{i}_{\alpha}. \tag{18}$$

If the rank p of the matrix $\| c_{ab} \|$ is equal to r, we have the theorem:

When a V_n admits p fields of parallel unit vectors, none of which is a null vector field orthogonal to all the other fields, each field is expressible linearly and homogeneously with constant coefficients in terms of p mutually orthogonal unit vector fields; and a coordinate system exists with respect to which the components of these unit vectors are given by (18) *and the fundamental quadratic form of the V_n is given by* (17).

2. When the fundamental form is definite, there being no possibility of real null vector fields, the above theorem gives the canonical form of the fundamental form of a V_n admitting p independent parallel vector fields. However, when the form is indefinite, there is the possibility of p independent unit fields and $r - p$ null fields, each field being orthogonal to the other fields of both sets. In this case we have to consider in addition to the vectors of components (5) those also of components (6). When the latter expressions are substituted in (1), on proceeding in a manner similar to that which led to equations (7), we obtain the equations

$$g_{\mu k}\frac{\partial \log \varphi_{\mu}^{2}}{\partial x^{j}} + \frac{\partial g_{\mu k}}{\partial x^{j}} + \frac{\partial g_{jk}}{\partial x^{\mu}} - \frac{\partial g_{\mu j}}{\partial x^{k}} = 0 \quad \begin{pmatrix} \mu = p+1, \cdots, r; \\ j, k = 1, \cdots, n \end{pmatrix}. \tag{19}$$

For $j = \alpha$ $(=1, \cdots, p)$ and $k = \sigma (= r+1, \cdots, n)$ the equations reduce to

$$g_{\mu\sigma}\frac{\partial \log \varphi_{\mu}^{2}}{\partial x^{\alpha}} = 0,$$

since $g_{\mu\sigma}$ are independent of $x^{1}, \cdots, x^{p}$ and $g_{\alpha\sigma} = g_{\alpha\mu} = 0$. Because of (4) and the understanding that the determinant of the g's is not equal to zero, we cannot have $g_{\mu\sigma} = 0$ for $\sigma = r+1, \cdots, n$. Hence the quantities φ_{μ} are independent of $x^{1}, \cdots, x^{p}$. Consequently the problem of these null vector fields is the problem of finding the form of the functions $g_{\sigma\tau}(x^{p+1}, \cdots, x^{n})$ for the form $g_{\sigma\tau}\, dx^{\sigma}\, dx^{\tau}$, and this is a problem independent of that stated in the above theorem.

Accordingly we consider the question of a V_n which admits q mutually orthogonal null parallel fields. As before there exists a coordinate system in

terms of which the components of the vectors have the values (6), where now $\mu = 1, \cdots, q$. Also we have equations (19) for $\mu = 1, \cdots, q$. Since we have

$$g_{\mu\nu} = 0 \qquad (\mu, \nu = 1, \cdots, q), \tag{20}$$

equations (19) for $k = \nu$ reduce to

$$\frac{\partial g_{\sigma\nu}}{\partial x^{\mu}} - \frac{\partial g_{\sigma\mu}}{\partial x^{\nu}} = 0 \qquad (\mu, \nu = 1, \cdots, q; \sigma = q+1, \cdots, n). \tag{21}$$

Also equations (19) for $j = \nu$, $k = \sigma$ reduce in consequence of (21) to

$$\frac{\partial}{\partial x^{\nu}} (\varphi_{\mu} g_{\mu\sigma}) = 0. \tag{22}$$

If from equation (19) for $j = \sigma$, $k = \tau$, for $\sigma, \tau = q+1, \cdots, n$, we subtract the equation for $j = \tau$, $k = \sigma$, we obtain

$$\frac{\partial}{\partial x^{\sigma}} (\varphi_{\mu} g_{\mu\tau}) = \frac{\partial}{\partial x^{\tau}} (\varphi_{\mu} g_{\mu\sigma}).$$

In consequence of this result and (22), we have

$$\varphi_{\mu} g_{\mu\sigma} = \frac{\partial \theta_{\mu}}{\partial x^{\sigma}}, \tag{23}$$

where θ_{μ} are functions of $x^{q+1}, \cdots, x^{n}$ at most. As a result of (23) we have that the functions θ_{μ} must be independent, otherwise the determinant $|g_{ij}|$ is equal to zero. Consequently $q \leqq n - q$, that is $q \leqq n/2$.

From (23) and (21) we have

$$\frac{\partial}{\partial x^{\nu}} \left(\frac{1}{\varphi_{\mu}}\right) \frac{\partial \theta_{\mu}}{\partial x^{\sigma}} = \frac{\partial}{\partial x^{\mu}} \left(\frac{1}{\varphi_{\nu}}\right) \frac{\partial \theta_{\nu}}{\partial x^{\sigma}}. \tag{24}$$

Since these equations must hold for $\sigma = q+1, \cdots, n$ and the rank of the jacobian matrix $\left\| \frac{\partial \theta_{\mu}}{\partial x^{\sigma}} \right\|$ is q, we have

$$\frac{\partial}{\partial x^{\nu}} \left(\frac{1}{\varphi_{\mu}}\right) = 0 \qquad (\mu, \nu = 1, \cdots, q; \mu \neq \nu).$$

Hence φ_{μ} is a function of $x^{\mu}, x^{q+1}, \cdots, x^{n}$ at most. If then we effect the non-singular transformation

$$x'^{\mu} = \int \frac{dx^{\mu}}{\varphi_{\mu}}, \qquad x'^{\sigma} = x^{\sigma} \quad (\mu = 1, \cdots, q; \sigma = q+1, \cdots, n),$$

the components of the vectors are

$$\lambda_{\mu}^{i} = \delta_{\mu}^{i}. \tag{25}$$

Also in this coordinate system we have equations (19) with $\varphi_\mu = 1$, from which equations we have (22) and (23) with $\varphi_\mu = 1$ and

$$\frac{\partial g_{\sigma\tau}}{\partial x^\mu} = 0. \tag{26}$$

We have remarked that the jacobian matrix $\left\| \frac{\partial \theta_\mu}{\partial x^\sigma} \right\|$ is of rank q. Without loss of generality we assume that the jacobian $\left| \frac{\partial \theta_\mu}{\partial x^\rho} \right|$, for $\rho = q+1, \cdots, 2q$, is not equal to zero. Accordingly a non-singular transformation is defined by

$$x'^h = x^h, \qquad x'^{q+\mu} = \theta_\mu \qquad \begin{pmatrix} h = 1, \cdots, q, 2q+1, \cdots, n; \\ \mu = 1, \cdots, q \end{pmatrix}.$$

In this new coordinate system the fundamental form is (dropping primes)

$$2\sum_\mu dx^\mu\, dx^{q+\mu} + g_{\sigma\tau}\, dx^\sigma\, dx^\tau \qquad \begin{pmatrix} \mu = 1, \cdots, q; \\ \sigma, \tau = q+1, \cdots, n \end{pmatrix}, \tag{27}$$

where $g_{\sigma\tau}$ are independent of $x^{q+1}, \cdots, x^n$, and the components of the vector fields are given by (25). Hence we have the theorem:

A V_n with indefinite fundamental form can admit at most $n/2$ mutually orthogonal fields of parallel null vectors; when it admits q such fields, the fundamental form is reducible to (27).

Gathering together the foregoing results we have the theorem:

When a V_n admits $p+q$ independent fields of parallel vectors such that p of the fields are expressible linearly, with constant coefficients, in terms of p mutually orthogonal unit fields, and the q fields consist of null vectors, each field being orthogonal to the other $p+q-1$ fields, there exists a coordinate system in terms of which the fundamental form is

$$\sum_\alpha e_\alpha (dx^\alpha)^2 + 2\sum_\mu dx^{p+\mu}\, dx^{p+q+\mu} + g_{\sigma\tau}\, dx^\sigma\, dx^\tau \qquad \begin{pmatrix} \alpha = 1, \cdots, p; \mu = 1, \cdots, q; \\ \sigma, \tau = p+q+1, \cdots, n \end{pmatrix},$$

where $g_{\sigma\tau}$ are functions of $x^{p+q+1}, \cdots, x^n$, at most, and the components of the fields are δ^i_ρ for $\rho = 1, \cdots, p+q$.

PRINCETON UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ÜBER DIE DICHTE DER SUMME VON MENGEN POSITIVER GANZER ZAHLEN. I

VON ALFRED BRAUER

(Received April 14, 1937; revised July 7, 1937)

Einleitung. Im Anschluß an die Schnirelmannschen Arbeiten über das Goldbachsche Problem ist die Frage viel untersucht worden, die Dichte der Summe mehrerer Mengen positiver ganzer Zahlen (in gleich näher zu definierendem Sinne) durch die Dichten der einzelnen Mengen abzuschätzen. Die vorliegende Arbeit soll einen Beitrag zu dieser Frage liefern und insbesondere Resultate der Herren E. Landau, A. Khintchine und I. Schur verbessern.

Die im folgenden betrachteten Mengen seien sämtlich Mengen positiver ganzer Zahlen. Ist A eine solche Menge, so werde für jede positive ganze Zahl x unter $A(x)$ die Anzahl der Elemente a aus A verstanden, für die $a \leqq x$ ist.

Herr L. Schnirelmann[1] hat die folgenden Definitionen eingeführt:

DEFINITION 1: *Unter der Summe C der n Mengen $A_1, A_2, \cdots, A_n$*

$$C = A_1 + A_2 + \cdots + A_n$$

verstehe man die Gesamtheit der positiven Zahlen c, die sich in der Form

$$c = e_1 a_1 + e_2 a_2 + \cdots + e_n a_n$$

darstellen lassen, wo a_ν für $\nu = 1, 2, \cdots, n$ ein Element von A_ν und $e_\nu = 0$ oder 1 ist.

Diese Summenbildung ist natürlich assoziativ und kommutativ.

DEFINITION 2: *Unter der Dichte α der Menge A verstehe man die untere Grenze der Quotienten $A(x)/x$ für $x = 1, 2, \cdots$.*

Es ist also $0 \leqq \alpha \leqq 1$. Dann und nur dann ist $\alpha = 1$, wenn A die Menge aller positiven ganzen Zahlen ist.

Es seien nun $A_1, A_2, \cdots, A_n$ Mengen mit den Dichten $\alpha_1, \alpha_2, \cdots, \alpha_n$. Ohne Beschränkung der Allgemeinheit kann man

$$\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n \tag{1}$$

annehmen. Für die Dichte γ der Summe $C = A_1 + A_2 + \cdots + A_n$ sollen möglichst gute untere Schranken $\phi(\alpha_1, \alpha_2, \cdots, \alpha_n)$ bestimmt werden, derart daß bei jeder Wahl der Mengen $A_1, A_2, \cdots, A_n$

$$\gamma \geqq \text{Min}\,\{1, \phi(\alpha_1, \alpha_2, \cdots, \alpha_n)\} \tag{2}$$

gilt.

[1] Ob additiwnich swoistwach tschissel (russisch), Iswestija Donskowo Polytechnitscheskowo Instituta (Nowotscherkask), Bd. 14 (1930), S. 3-28, und Über additive Eigenschaften der Zahlen, Math. Annalen, Bd. 107 (1933), S. 649-690.

Ist $\gamma = 1$, so ist (2) stets erfüllt. Um daher für eine Funktion

$$\phi(\alpha_1, \alpha_2, \cdots, \alpha_n)$$

zu beweisen, daß sie eine Schranke im Sinne von (2) liefert, genügt es zu zeigen, daß

$$\gamma \geqq \phi(\alpha_1, \alpha_2, \cdots, \alpha_n) \quad \text{für} \quad \gamma < 1 \tag{3}$$

ist. Speziell folgt aus (2) und daher auch aus (3)

$$\gamma = 1 \quad \text{für} \quad \phi(\alpha_1, \alpha_2, \cdots, \alpha_n) \geqq 1,$$

da γ als Dichte einer Menge nicht größer als 1 sein kann.

Herr Schnirelmann zeigte nun: *Sind A_1 und A_2 zwei Mengen mit den Dichten α_1 und α_2 und ist $\alpha_1 + \alpha_2 \geqq 1$, so besitzt $A_1 + A_2$ die Dichte $\gamma = 1$.*

Für die Funktion $\phi(\alpha_1, \alpha_2) = \alpha_1 + \alpha_2$ ist also (2) erfüllt, falls $\alpha_1 + \alpha_2 \geqq 1$ ist. Im Anschluß an dieses Resultat ist oft die Vermutung ausgesprochen worden, daß $\alpha_1 + \alpha_2$ der Ungleichung (2) stets genügt, daß also

$$\gamma \geqq \alpha_1 + \alpha_2 \quad \text{für} \quad \gamma < 1 \tag{4}$$

gilt. Wäre dies richtig, so würde sofort für beliebiges $n > 2$

$$\gamma \geqq \alpha_1 + \alpha_2 + \cdots + \alpha_n \quad \text{für} \quad \gamma < 1 \tag{4a}$$

folgen. Bisher ist es nicht gelungen (4) und (4a) zu beweisen. Man kennt nur die folgenden Teilresultate:

I. (E. Landau.[2]) *Für $n = 2$ ist $\gamma \geqq \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.*

II. (A. Khintchine.[3]) *Ist (1) erfüllt und $\gamma < 1$, so gilt $\gamma \geqq n\alpha_1$ für jedes n.*

Aus dem Satz II folgt die Richtigkeit von (4) und (4a) für den Spezialfall

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n . \tag{5}$$

III. (A. Besicovitch.[4]) *Es seien A und B zwei Mengen mit den Dichten α und β, ferner sei α^* die untere Grenze der Quotienten $A(x)/(x+1)$ für $x = 1, 2, \cdots$. Hat dann $A + B$ die Dichte $\gamma < 1$, so ist $\gamma \geqq \alpha^* + \beta$.*

IV. (I. Schur.[5]) Für $n = 2$, $\alpha_1 \leqq \alpha_2$ und $\gamma < 1$ ist

$$\gamma \geqq \alpha_2/(1 - \alpha_1) \geqq \alpha_1/(1 - \alpha_2).$$

Hierin ist der Schnirelmannsche Satz als Spezialfall enthalten.

V. (I. Schur.[6]) *Für $n = 2$, $\alpha_1 \leqq \alpha_2$ und $\gamma < 1$ ist*

$$\gamma \geqq \tfrac{1}{2}\alpha_1 + \tfrac{1}{2}(\alpha_1^2 + 4\alpha_2^2)^{\frac{1}{2}} \geqq \tfrac{1}{2}\alpha_2 + \tfrac{1}{2}(\alpha_2^2 + 4\alpha_1^2)^{\frac{1}{2}}.$$

---

[2] Die Goldbachsche Vermutung und der Schnirelmannsche Satz, Göttinger Nachrichten, Math.-Phys. Klasse, 1930, S. 255–276.

[3] Zur additiven Zahlentheorie, Matematitscheski Sbornik, Bd. 39 Heft 3 (1932), S. 27–34.

[4] On the density of the sum of two sequences of integers, Journal of the London Math. Society, Bd. 10 (1935), S. 246–248.

[5] Über den Begriff der Dichte in der additiven Zahlentheorie, Sitzungsberichte der Preußischen Akademie der Wissenschaften, Phys.-Math. Klasse, 1936, S. 269–297.

[6] A. a. O. 5).

Ein Beweis der Sätze I und III und des Satzes II für $n = 2$ findet sich auch in einem demnächst erscheinenden Buche von Herrn E. Landau;[7] in diesem Buch wird auf die Vermutung (4) als ungelöstes Problem besonders aufmerksam gemacht.

Im folgenden zeige ich zunächst, daß aus dem Khintchineschen Beweis von II sofort die Richtigkeit von (4) in zwei weiteren Spezialfällen folgt. Aus diesem Beweis ergibt sich nämlich

SATZ 1: *Es seien* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten* $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ *mit* $\alpha_1 > 0$; *ferner sei* γ *die Dichte von* $A_1 + A_2 + \cdots + A_n$.

1) *Ist dann* $\alpha_1 < 1/n$ *und* $\alpha_n \geqq 1 - n\alpha_1$, *so ist* $\gamma \geqq 1 - \alpha_1$.

2) *Ist* $\alpha_2 < 1/n$ *und* $\alpha_1 \geqq 1 - n\alpha_2$, *so ist* $\gamma \geqq 1 - \alpha_2$.

Dieser Satz liefert die Richtigkeit von (4) und (4a) in den beiden Spezialfällen

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{n-1} < \frac{1}{n}, \qquad \alpha_n = 1 - n\alpha_1 \tag{6}$$

und

$$\alpha_2 = \alpha_3 = \cdots = \alpha_n < \frac{1}{n}, \qquad \alpha_1 = 1 - n\alpha_2. \tag{7}$$

Wählt man z. B. $n = 2$ und $\alpha_1 = \frac{1}{4}$, so folgt aus (5), (6) und (7) die Richtigkeit von (4) für $\alpha_2 = \frac{1}{4}$, $\alpha_2 = \frac{1}{2}$ und $\alpha_2 = \frac{3}{8}$. Für $n > 2$ erhält man durch Kombination von (5), (6) und (7) weitere Fälle, in denen (4a) gilt.

Für $\alpha_1 = 0$ oder $\alpha_2 = 0$ ist (4) trivial, da die Summenmenge die einzelnen Mengen enthält. Daher genügt es nach dem Obigen, $\alpha_1 > 0$, $\alpha_2 > 0$ und $\alpha_1 + \alpha_2 \leqq 1$ vorauszusetzen. Unter Benutzung des Satzes von Besicovitch (III) beweise ich den folgenden

SATZ 2: *Es seien* A_1 *und* A_2 *Mengen mit den Dichten* $\alpha_1 > 0$ *und* $\alpha_2 > 0$ *mit* $\alpha_1 + \alpha_2 \leqq 1$. *Die ganzen Zahlen* k *und* l *seien so bestimmt, daß*

$$\frac{1}{k-1} \geqq \alpha_1 > \frac{1}{k}, \qquad \frac{1}{l-1} \geqq \alpha_2 > \frac{1}{l}$$

ist. Ist γ *die Dichte der Menge* $A_1 + A_2$, *so ist*

$$\gamma \geqq \operatorname{Max}\left\{\alpha_2 + \frac{1}{k}, \alpha_1 + \frac{1}{l}\right\} = \operatorname{Max}\left\{\alpha_2 + \frac{1}{1 + [1/\alpha_1]}, \alpha_1 + \frac{1}{1 + [1/\alpha_2]}\right\}.$$

Aus diesem Satz folgt sofort der etwas weniger scharfe

SATZ 3: *Es ist für* $\alpha_1 \leqq \alpha_2$ *mit* $\alpha_1 + \alpha_2 \leqq 1$

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{1 + \alpha_1} \geqq \alpha_1 + \frac{\alpha_2}{1 + \alpha_2}.$$

---

[7] Über einige neuere Fortschritte der additiven Zahlentheorie, Cambridge (1937). Vgl. insbesondere S. 3-4, 7-8, 56-67 und 71-79. Zusatz bei der Korrektur: Dieses Buch ist inzwischen erschienen.

Sind mehr als zwei Mengen $A_1, A_2, \cdots, A_n$ mit den positiven Dichten $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ gegeben, so folgt hieraus, falls $A_1 + A_2 + \cdots + A_n$ die Dichte $\gamma < 1$ hat, für $\nu = 1, 2, \cdots, n$

$$\gamma \geqq \alpha_\nu + \sum_{\substack{\mu=1 \\ \mu \neq \nu}}^{n} \frac{1}{1 + [1/\alpha_\mu]} \geqq \alpha_\nu + \sum_{\substack{\mu=1 \\ \mu \neq \nu}}^{n} \frac{\alpha_\mu}{1 + \alpha_\mu}.$$

Die Sätze 2 und 3 liefern für alle Werte von α_1 und α_2, die den Bedingungen $0 < \alpha_1 \leqq \alpha_2$ und $\alpha_1 + \alpha_2 \leqq 1$ genügen, bessere Abschätzungen als der Landausche Satz I und der Schursche Satz V.

Vergleicht man die Sätze 1 und 2 mit dem Satz von Khintchine (II) und dem Schurschen Satz IV, so zeigt sich, wenn $\alpha_1 \leqq \alpha_2$ ist, daß für jeden Wert von α_1 mit $0 < \alpha_1 < \frac{1}{2}$ sowohl Werte von α_2 existieren, für die II, wie Werte, für die IV, wie Werte, für die Satz 2 eine bessere Abschätzung liefert als die übrigen drei dieser Sätze. Für $\alpha_1 < \frac{1}{3}$ existieren außerdem zu jedem Werte von α_1 Werte von α_2, für die Satz 1 eine bessere Abschätzung als die drei anderen dieser Sätze liefert. Um sofort zu übersehen, welche dieser Abschätzungen für gegebenes α_1, α_2 die beste ist, deute ich α_1, α_2 als rechtwinklige Koordinaten und bestimme die Gebiete der $\alpha_1\alpha_2$-Ebene, in denen jeder dieser Sätze die beste Abschätzung liefert (vgl. hierzu die Figur auf S. 331).

Durch Kombination der Sätze I, II, IV und V hat Herr Schur[8] eine Reihe weiterer Sätze über die Dichte der Summe von n Mengen abgeleitet, von denen hier zunächst die folgenden genannt seien:

VI. *Sind $A_1, A_2, \cdots, A_n$ Mengen mit den Dichten $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ und hat $A_1 + A_2 + \cdots + A_n$ die Dichte $\gamma < 1$, so ist*

$$\gamma \geqq \alpha_n + \frac{1}{2}(\alpha_{n-1} + \alpha_{n-2} + \cdots + \alpha_1).$$

VII. *$A_1, A_2, \cdots, A_n$ seien Mengen mit den Dichten $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$; es sei β die Dichte von $A_2 + A_3 + \cdots + A_n$ und $\gamma < 1$ die von*

$$A_1 + A_2 + \cdots + A_n.$$

Genügt die Zahl a der Bedingung

$$0 < a \leqq 2(n^2 - n)^{\frac{1}{2}}/(2n - 1) \tag{8}$$

so gilt $\gamma \geqq \alpha_1[n - a(n^2 - n)^{\frac{1}{2}}] + a\beta$. Speziell folgt für $n = 2$ und $a = 1/\sqrt{2}$.

VIIa. $\gamma \geqq \alpha_1 + \alpha_2/\sqrt{2}$.

VIII. *Für jedes $n \geqq 2$ sei c_n die größte Zahl, derart daß bei beliebiger Wahl der Mengen $A_1, A_2, \cdots, A_n$ mit den Dichten $\alpha_1, \alpha_2, \cdots, \alpha_n$ die Dichte γ von $A_1 + A_2 + \cdots + A_n$ für $\gamma < 1$ der Bedingung $\gamma \geqq c_n(\alpha_1 + \alpha_2 + \cdots + \alpha_n)$ genügt. Dann ist*

$$c_2 \geqq \frac{2}{1 + \sqrt{2}} > 0{,}8284, \qquad c_3 \geqq \frac{3\sqrt{2}}{\sqrt{2} + \sqrt{3} + \sqrt{6}} > 0{,}7581$$

[8] A. a. O. 5).

und für $n \geqq 4$

$$c_n \geqq \frac{1}{\frac{1}{c_3} + \sum_{\nu=4}^{\infty} \frac{1}{\sqrt{\nu^4 - \nu^3}}} > 0{,}6148.$$

Mittels des Satzes 2 werden diese Sätze im folgenden bedeutend verschärft und zugleich die Beweise gegenüber den Schurschen Beweisen vereinfacht.

Es ergibt sich statt VI

$$\gamma \geqq \alpha_n + \frac{3}{4}\alpha_{n-1} + \frac{4}{5}\alpha_{n-2} + \cdots + \frac{n+1}{n+2}\alpha_1.$$

Der Satz VII kann zu

$$\gamma \geqq \alpha_1\left(n - \frac{n^2+n-1}{n+2}a\right) + a\beta$$

verschärft werden, wobei zugleich (8) durch die schwächere Voraussetzung

$$0 < a < 1$$

ersetzt werden kann. Speziell ergibt sich statt VIIa

$$\gamma \geqq \alpha_1 + \frac{4}{5}\alpha_2.$$

Statt des Satzes VIII gilt

$$c_2 \geqq \frac{8}{9} > 0{,}8888; \qquad c_3 \geqq \frac{120}{139} > 0{,}8633;$$

$$c_n \geqq \frac{2880}{3361} > 0{,}8568 \quad \text{für} \quad n \geqq 4.$$

Auf die Verschärfung einiger weiterer Sätze von Herrn Schur soll in einer Fortsetzung zu dieser Arbeit eingegangen werden.

1. **Beweis der Hauptsätze.** Herr Khintchine erbringt den Beweis von II auf Grund von zwei Hilfssätzen, die er durch gleichzeitige Induktion beweist. Der zweite dieser Hilfssätze lautet folgendermaßen:

HILFSSATZ: *Es sei* $n \geqq 2$ *eine ganze Zahl; die Zahlen* α *und* μ *mögen der Bedingung*

(9) $$0 \leqq \mu < \alpha < \frac{1}{n}$$

genügen. Es sei S die Summe der n Mengen $A_1, A_2, \cdots, A_n$. Ist dann für $x = 1, 2, \cdots, X$

$$A_1(x) \geqq \alpha x,$$

$$A_2(x) \geqq \alpha x - \frac{1}{n},$$

$$A_3(x) \geqq \alpha x - \frac{2}{n}, \tag{10}$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$A_{n-1}(x) \geqq \alpha x - \frac{n-2}{n}$$

und

$$A_n(x) \geqq (1 - n\alpha)x - n\mu, \tag{11}$$

so ist

$$S(X) \geqq (1 - \alpha)X - \mu. \tag{12}$$

Aus diesem Hilfssatz ergibt sich nun sofort

SATZ 1: *Es seien $A_1, A_2, \cdots, A_n$ Mengen mit den positiven Dichten*

$$\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n. \tag{13}$$

Die Dichte von $A_1 + A_2 + \cdots + A_n$ sei γ.

1) *Genügen dann α_1 und α_n den beiden Bedingungen*

$$\alpha_1 < \frac{1}{n}, \tag{14}$$

$$\alpha_n \geqq 1 - n\alpha_1, \tag{15}$$

so ist

$$\gamma \geqq 1 - \alpha_1. \tag{16}$$

2) *Genügen α_1 und α_2 den beiden Bedingungen*

$$\alpha_2 < \frac{1}{n}, \tag{17}$$

$$\alpha_1 \geqq 1 - n\alpha_2, \tag{18}$$

so ist

$$\gamma \geqq 1 - \alpha_2. \tag{19}$$

BEWEIS: 1) Man setze $\alpha_1 = \alpha$ und wende den Hilfssatz für $\mu = 0$ auf die Mengen $A_1, A_2, \cdots, A_n$ an. Die Voraussetzungen dieses Hilfssatzes sind erfüllt; denn aus (14) folgt (9). Aus (13) folgt für jedes positive ganzzahlige x

$$A_\nu(x) \geqq \alpha_\nu x \geqq \alpha_1 x \geqq \alpha_1 x - \frac{\nu - 1}{n} \qquad (\nu = 1, 2, \cdots, n-1);$$

also ist (10) erfüllt. Ferner folgt (11) aus (15). Daher gilt (12) für jedes positive ganzzahlige X; dies liefert (16).

2) Man setze $\alpha_2 = \alpha$ und wende den Hilfssatz für $\mu = 0$ auf die Mengen $A_2, A_3, \cdots, A_n, A_1$ an. (9) folgt aus (17) und (11) aus (18). Aus (13) folgt

$$A_\nu(x) \geqq \alpha_\nu x \geqq \alpha_2 x \geqq \alpha_2 x - \frac{\nu - 2}{n} \qquad (\nu = 2, 3, \cdots, n);$$

also ist auch (10) erfüllt, und (12) liefert (19).

Ist speziell

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{n-1} < \frac{1}{n},$$

$$\alpha_n = 1 - n\alpha_1,$$

so sind die Voraussetzungen des Satzes 1 erfüllt, und man erhält

$$\gamma \geqq 1 - \alpha_1 = (n-1)\alpha_1 + 1 - n\alpha_1 = \alpha_1 + \alpha_2 + \cdots + \alpha_{n-1} + \alpha_n.$$

In diesem Falle ist also die Vermutung (4a) bewiesen. Analog folgt dies für den Spezialfall

$$\alpha_2 = \alpha_3 = \cdots = \alpha_n < \frac{1}{n},$$

$$\alpha_1 = 1 - n\alpha_2.$$

SATZ 2: *Es seien A_1 und A_2 zwei Mengen mit den Dichten $\alpha_1 > 0$ und $\alpha_2 > 0$ mit $\alpha_1 + \alpha_2 \leqq 1$. Die ganzen Zahlen k und l seien so bestimmt, daß*

$$\frac{1}{k-1} \geqq \alpha_1 > \frac{1}{k}, \qquad \frac{1}{l-1} \geqq \alpha_2 > \frac{1}{l} \tag{20}$$

ist. Ist γ die Dichte der Menge $A_1 + A_2$, so ist

$$\gamma \geqq \operatorname{Max}\left\{\alpha_2 + \frac{1}{k}, \alpha_1 + \frac{1}{l}\right\} = \operatorname{Max}\left\{\alpha_2 + \frac{1}{1 + [1/\alpha_1]}, \alpha_1 + \frac{1}{1 + [1/\alpha_2]}\right\}.$$

BEWEIS: Es sei α_1^* die untere Grenze der Quotienten $A_1(x)/(x+1)$ für positives ganzzahliges x; dann ist nach III

$$\gamma \geqq \alpha_2 + \alpha_1^*. \tag{21}$$

Man dividiere nun für beliebiges positives ganzzahliges x die Zahl x durch k

$$x = gk + r \qquad (0 \leqq r < k). \tag{22}$$

Es sei nun zunächst $g > 0$. Wegen (20) ist $\alpha_1 k > 1$, daher ist

$$A_1(x) \geqq \alpha_1 x = \alpha_1 gk + \alpha_1 r > g + \alpha_1 r.$$

Nun ist $A_1(x)$ eine ganze Zahl; da g ganz ist, ist daher $A_1(x) \geqq g + 1$. Für $g = 0$ gilt diese Abschätzung ebenfalls, da $x \geqq 1$ war.

Wegen (22) ist $r + 1 \leqq k$; folglich ist für $x \geqq 1$

$$\frac{A_1(x)}{x+1} \geqq \frac{g+1}{gk+r+1} \geqq \frac{g+1}{gk+k} = \frac{1}{k},$$

also

$$\alpha_1^* \geqq \frac{1}{k}$$

und wegen (21)

$$\gamma \geqq \alpha_2 + \frac{1}{k}. \tag{23}$$

Da α_1 und α_2 denselben Bedingungen genügen, erhält man analog

$$\gamma \geqq \alpha_1 + \frac{1}{l}. \tag{24}$$

Aus (20) folgt

$$k - 1 \leqq \frac{1}{\alpha_1} < k.$$

Daher ist

$$k = \left[\frac{1}{\alpha_1}\right] + 1 \tag{25}$$

und analog

$$l = \left[\frac{1}{\alpha_2}\right] + 1. \tag{26}$$

Aus (23), (24), (25) und (26) folgt

$$\gamma \geqq \operatorname{Max}\left\{\alpha_2 + \frac{1}{k}, \alpha_1 + \frac{1}{l}\right\} = \operatorname{Max}\left\{\alpha_2 + \frac{1}{1+[1/\alpha_1]}, \alpha_1 + \frac{1}{1+[1/\alpha_2]}\right\}.$$

Etwas weniger scharf läßt sich Satz 2 folgendermaßen formulieren:

SATZ 3: *Genügen die Dichten α_1 und α_2 der Mengen A_1 und A_2 den Bedingungen $\alpha_1 \leqq \alpha_2$ und $\alpha_1 + \alpha_2 \leqq 1$, so ist die Dichte γ der Summe $A_1 + A_2$*

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{1+\alpha_1} \geqq \alpha_1 + \frac{\alpha_2}{1+\alpha_2}.$$

BEWEIS: Für $\alpha_1 = 0$ ist die Behauptung trivial, für $\alpha_1 > 0$ folgt sie aus

$$\alpha_2 + \frac{1}{1+[1/\alpha_1]} \geqq \alpha_2 + \frac{\alpha_1}{1+\alpha_1}.$$

Der Satz 3, und daher erst recht Satz 2, liefert eine bessere Abschätzung als der Satz I, denn es ist für $\alpha_1 > 0$

$$\alpha_2 + \frac{\alpha_1}{1+\alpha_1} = \alpha_2 + \alpha_1 - \frac{\alpha_1^2}{1+\alpha_1} > \alpha_2 + \alpha_1 - \alpha_1^2 \geqq \alpha_2 + \alpha_1 - \alpha_1\alpha_2.$$

Ferner soll gezeigt werden, daß Satz 3 auch eine bessere Abschätzung als der Satz V liefert. Wegen $\alpha_1 \leqq \alpha_2$ kann $\alpha_1 \leqq \frac{1}{2}$ angenommen werden; daher ist

$$\alpha_2 + \frac{\alpha_1}{1+\alpha_1} \geqq \alpha_2 + \frac{\alpha_1}{1+\frac{1}{2}} = \alpha_2 + \frac{2}{3}\alpha_1 = \frac{\alpha_1}{2} + \alpha_2 + \frac{\alpha_1}{6}$$

$$= \frac{\alpha_1}{2} + \frac{1}{2}\left[\left(2\alpha_2 + \frac{\alpha_1}{3}\right)^2\right]^{\frac{1}{2}} = \frac{\alpha_1}{2} + \frac{1}{2}\left(4\alpha_2^2 + \frac{4}{3}\alpha_2\alpha_1 + \frac{\alpha_1^2}{9}\right)^{\frac{1}{2}} \tag{27}$$

$$\geqq \frac{\alpha_1}{2} + \frac{1}{2}\left(4\alpha_2^2 + \frac{4}{3}\alpha_1^2 + \frac{\alpha_1^2}{9}\right)^{\frac{1}{2}} > \frac{\alpha_1}{2} + \frac{1}{2}(4\alpha_2^2 + \alpha_1^2)^{\frac{1}{2}}.$$

Aus (27) folgt ferner

SATZ 4: *Ist* $\alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 \leqq 1$, *so ist*

$$\gamma \geqq \alpha_2 + \frac{2}{3}\alpha_1.$$

Aus den Sätzen II, IV, 1 und 2 folgt

SATZ 5: *Ist* $0 < \alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 \leqq 1$, *so ist*

$$\gamma \geqq \operatorname{Max}\left\{2\alpha_1,\ \frac{\alpha_2}{1-\alpha_1},\ \alpha_2 + \frac{1}{1+[1/\alpha_1]},\ \alpha_1 + \frac{1}{1+[1/\alpha_2]}\right\} = M.$$

Ferner ist

$$\gamma \geqq \operatorname{Max}\{M, 1-\alpha_2\} \quad \textit{für} \quad \alpha_1 + 2\alpha_2 \geqq 1,$$

$$\gamma \geqq \operatorname{Max}\{M, 1-\alpha_1\} \quad \textit{für} \quad 2\alpha_1 + \alpha_2 \geqq 1.$$

Da die Addition der Mengen assoziativ ist, lassen sich die Sätze 2 und 3 sofort auch für mehr als zwei Mengen aussprechen. Es gilt

SATZ 6: *Sind* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten*

$$0 < \alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$$

und hat die Summe $A_1 + A_2 + \cdots + A_n$ *die Dichte* $\gamma < 1$, *so ist für*

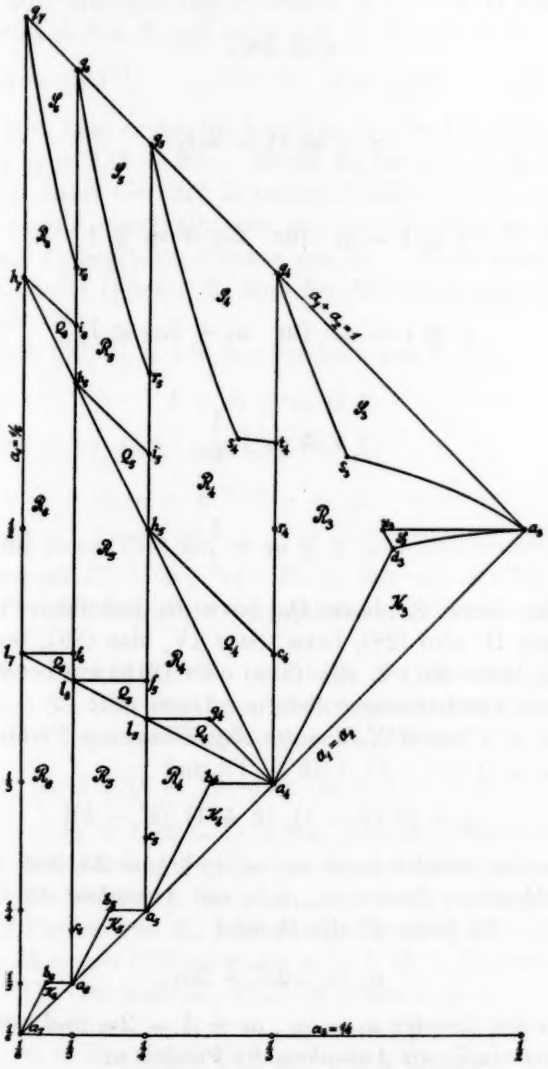
$$\nu = 1, 2, \cdots, n$$

$$\gamma \geqq \alpha_\nu + \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} \frac{1}{1+[1/\alpha_\mu]} \geqq \alpha_\nu + \sum_{\substack{\mu=1 \\ \mu\neq\nu}}^{n} \frac{\alpha_\mu}{1+\alpha_\mu}.$$

2. **Vergleich der Sätze II, IV, 1 und 2.** Es soll nun untersucht werden, wann für $n = 2$ und gegebenes α_1 und α_2 der Satz II, wann Satz IV, wann Satz 1 und wann Satz 2 eine bessere Abschätzung liefert als die übrigen drei dieser Sätze. Zu diesem Zwecke deuten wir α_1 und α_2 als rechtwinklige Koordinaten in der Ebene. Wegen $0 \leqq \alpha_1 \leqq \alpha_2$ und $\alpha_1 + \alpha_2 \leqq 1$ genügt es, das abgeschlossene

Dreieck D zu betrachten, das von den Geraden $\alpha_1 = 0$, $\alpha_1 = \alpha_2$ und $\alpha_1 + \alpha_2 = 1$ begrenzt wird. Es sei G_k der Durchschnitt von D mit dem Streifen

$$\frac{1}{k} < \alpha_1 \leqq \frac{1}{k-1} \qquad (k = 3, 4, \ldots).$$



In der Figur auf Seite . . . sind G_3, G_4, G_5 und G_6 dargestellt.

Auf den Seiten von D gilt

$$\gamma \geqq \alpha_1 + \alpha_2,$$

da dies für $\alpha_1 = 0$ trivial ist und für $\alpha_1 = \alpha_2$ aus II, für $\alpha_1 + \alpha_2 = 1$ aus IV folgt. Abgesehen von diesen drei Geraden ist die Vermutung (4) durch die

Spezialfälle des Satzes 1 auf den beiden Geraden $\alpha_1 + 2\alpha_2 = 1$, $2\alpha_1 + \alpha_2 = 1$ bewiesen. Diese Geraden schneiden die Gerade $\alpha_1 = \alpha_2$ im Punkte a_4 mit den Koordinaten $(\frac{1}{3}, \frac{1}{3})$ und die Gerade $\alpha_1 = 1/(k-1)$ für $k \geqq 5$ in den Punkten $l_k = [1/(k-1), (k-2)/2(k-1)]$ und $h_k = [1/(k-1), (k-3)/(k-1)]$.

In G_k liefert Satz II

$$\gamma \geqq 2\alpha_1 , \tag{28}$$

Satz IV

$$\gamma \geqq \alpha_2/(1-\alpha_1), \tag{29}$$

Satz 1

$$\gamma \geqq 1 - \alpha_1 \quad \text{für} \quad 2\alpha_1 + \alpha_2 \geqq 1 \tag{30a}$$

und

$$\gamma \geqq 1 - \alpha_2 \quad \text{für} \quad \alpha_1 + 2\alpha_2 \geqq 1, \tag{30b}$$

Satz 2

$$\gamma \geqq \alpha_2 + \frac{1}{k} \tag{31a}$$

und

$$\gamma \geqq \alpha_1 + \frac{1}{l} \qquad \text{für } l = [1/\alpha_2] + 1. \tag{31b}$$

Es seien nun K_k, bezw. S_k, bezw. Q_k, bezw. R_k diejenigen Punktmengen von G_k, in denen Satz II, also (28), bezw. Satz IV, also (29), bezw. Satz 1, also (30a) oder (30b), bezw. Satz 2, also (31a) oder (31b) ein besseres Resultat als die übrigen dieser Abschätzungen liefern. Dann gilt:

SATZ 7: *Für $k \geqq 4$ besteht K_k aus der abgeschlossenen Verbindungsstrecke der beiden Punkte $a_k = [1/(k-1), 1/(k-1)]$ und*

$$c_k = [1/(k-1), (k+1)/(k^2-k)]$$

und aus dem von den Geraden $\alpha_1 = \alpha_2$, $\alpha_2 + 1/k = 2\alpha_1$ und $\alpha_2 = 1/(k-1)$ gebildeten abgeschlossenen Dreieck $a_{k+1}a_kb_k$ mit Ausnahme der nicht zu G_k gehörenden Ecke a_{k+1}. Ist ferner P_1 die Parabel

$$\alpha_2 = -2\alpha_1^2 + 2\alpha_1 , \tag{32}$$

so ist K_3 das von den Geraden $\alpha_1 = \alpha_2$, $\alpha_2 + \frac{1}{3} = 2\alpha_1$ und von P_1 begrenzte abgeschlossene Gebiet $a_4a_3d_3$ mit Ausnahme des Punktes a_4.

BEWEIS: Es ist (29) schärfer als (28), wenn $\alpha_2/(1-\alpha_1) > 2\alpha_1$, also

$$\alpha_2 > 2\alpha_1 - 2\alpha_1^2 \tag{33}$$

ist. Wegen (32) liefert daher oberhalb von P_1 der Satz IV und unterhalb dieser Kurve der Satz II eine bessere Abschätzung, auf P_1 ergeben beide Sätze dasselbe Resultat.

Ferner ist (31a) besser als (28), wenn

$$\alpha_2 + 1/k > 2\alpha_1 \tag{34}$$

ist. Daher gehören diejenigen Punkte von G_k, die oberhalb der Geraden $\alpha_2 + 1/k = 2\alpha_1$ oder oberhalb von P_1 liegen, nicht zu K_k.

Der Schnittpunkt d_k von P_1 mit $\alpha_2 + 1/k = 2\alpha_1$ hat die Koordinaten

$$\alpha_1 = (2k)^{-\frac{1}{2}}, \qquad \alpha_2 = [(2k)^{\frac{1}{2}} - 1]/k. \tag{35}$$

Wie man leicht sieht, liegt er nur für $k = 3$ in G_k; für $k \geqq 4$ liegt P_1 in G_k oberhalb der Geraden $\alpha_2 + 1/k = 2\alpha_1$. Es sei K_k' für $k \geqq 4$ die Gesamtheit der unterhalb oder auf dieser Geraden liegenden Punkte von G_k und für $k = 3$ die Gesamtheit der sowohl unterhalb oder auf der Geraden $\alpha_2 + \frac{1}{3} = 2\alpha_1$, wie unterhalb oder auf P_1 liegenden Punkte von G_k. Es ist nun noch zu prüfen, in welchen Punkten der Gebiete K_k' eine der Abschätzungen (31b), (30a) oder (30b) besser ist als (28).

In K_3' ist, da nach (32) $\alpha_2 \leqq \frac{1}{2}$ in den Punkten von P_1 ist,

$$\tfrac{1}{3} < \alpha_1 \leqq \alpha_2 \leqq \tfrac{1}{2}. \tag{36}$$

Wegen (36) liefern (31b), (30a) und (30b) in K_3'

$$\gamma \geqq \alpha_1 + \tfrac{1}{3} > 1 - \alpha_1 \geqq 1 - \alpha_2 .$$

Nun ist aber in K_3' wegen (36) $2\alpha_1 > \alpha_1 + \frac{1}{3}$, also (28) besser als (31b), (30a) und (30b). Daher ist $K_3' = K_3$, und K_3 ist das von den Geraden $\alpha_1 = \alpha_2$, $\alpha_2 + \frac{1}{3} = 2\alpha_1$ und von P_1 begrenzte abgeschlossene Gebiet $a_4 a_3 d_3$ mit Ausnahme des Punktes a_4.

Nun sei $k \geqq 4$. Es war K_k' das unterhalb der Geraden $\alpha_2 + 1/k = 2\alpha_1$ liegende Gebiet von G_k, also das abgeschlossene Dreieck $a_{k+1} a_k c_k$ mit Ausnahme der Ecke a_{k+1}. Für diejenigen Punkte von K_k', für die

$$\alpha_1 < 1/(k-1) \quad \text{und} \quad \alpha_2 > 1/(k-1) \tag{37}$$

gilt, liefert (31b) $\gamma \geqq \alpha_1 + 1/(k-1) > 2\alpha_1$, also eine bessere Abschätzung als (28). Die durch (37) bestimmten Punkte des Dreiecks $a_{k+1} a_k c_k$ gehören also nicht zu K_k.

Für die übrigen Punkte von K_k' muß man noch (30a) und (30b) betrachten. Für diese Punkte ist wegen (37) entweder $\alpha_1 = 1/(k-1)$ oder $\alpha_1 < 1/(k-1)$ und $\alpha_2 \leqq 1/(k-1)$. Im letzteren Falle ist wegen $k \geqq 4$

$$\alpha_1 + 2\alpha_2 < \frac{1}{k-1} + \frac{2}{k-1} = \frac{3}{k-1} \leqq 1,$$

also $\alpha_1 + 2\alpha_2 < 1$, folglich (30a) und (30b) nicht anwendbar.

Ist aber $\alpha_1 = 1/(k-1)$, so ist, da die Ecke c_k von K_k' die Koordinaten $[1/(k-1), (k+1)/(k^2-k)]$ hat, in K_k'

$$\alpha_2 \leqq \frac{k+1}{k(k-1)}.$$

Es sei nun zunächst $k \geqq 5$. Dann ist

$$\alpha_1 + 2\alpha_2 \leqq \frac{1}{k-1} + \frac{2k+2}{k(k-1)} = \frac{3k+2}{k(k-1)} \leqq \frac{17}{20} < 1.$$

Für diese Punkte ist also (30a) und (30b) ebenfalls nicht anwendbar.

Für $k = 4$, also für die Punkte der Strecke a_4c_4 ist

$$\alpha_1 = 1/(k-1) = \tfrac{1}{3}.$$

Für diese Punkte liefern (30a) und (28) dasselbe Resultat

$$\gamma \geqq 2\alpha_1 = 1 - \alpha_1,$$

und (30b) ist niemals besser als (30a). Die Punkte der Strecke a_4c_4 gehören also sowohl zu K_4 wie zu Q_4.

Daher gehören alle Punkte von K_k', für die (37) nicht gilt, zu K_k. Es sind dies einerseits die Punkte der Strecke a_kc_k, für die $\alpha_1 = 1/(k-1)$ ist, und andererseits die Punkte des von den Geraden $\alpha_1 = \alpha_2$, $\alpha_2 + 1/k = 2\alpha_1$ und $\alpha_2 = 1/(k-1)$ begrenzten abgeschlossenen Dreiecks $a_{k+1}a_kb_k$ mit Ausnahme der Ecke a_{k+1}; für diese Punkte ist $\alpha_2 \leqq 1/(k-1)$.

Damit ist Satz 7 bewiesen. In analoger Weise ergeben sich die folgenden drei Sätze, deren Beweis übergangen werde:

Satz 8: *Bezeichnet H_k die Hyperbel*

$$k\alpha_1\alpha_2 + \alpha_1 = 1 \qquad (k = 3, 4, \cdots),$$

so ist S_k für $k \geqq 5$ das von H_k und den Geraden $\alpha_1 = 1/(k-1)$ und $\alpha_1 + \alpha_2 = 1$ begrenzte abgeschlossene Gebiet $q_{k+1}q_kr_k$ mit Ausnahme des Punktes q_{k+1}. Ist ferner P_2 die Parabel

$$\alpha_2 = -\alpha_1^2 + \frac{\alpha_1}{2} + \frac{1}{2},$$

so besteht S_4 aus dem isolierten Punkt r_4 mit den Koordinaten $(\frac{1}{3}, \frac{1}{2})$ und aus dem von H_k, P_2 und den Geraden $\alpha_1 = \frac{1}{3}$ und $\alpha_1 + \alpha_2 = 1$ begrenzten abgeschlossenen Gebiet $q_5q_4t_4s_4$ mit Ausnahme des Punktes q_5. Endlich besteht S_3 aus den beiden abgeschlossenen Gebieten $q_4a_3s_3$ und $a_3d_3u_3$ mit Ausnahme des Punktes q_4; der Rand dieser Gebiete wird von H_3, P_2 und $\alpha_1 + \alpha_2 = 1$ bezw. von H_3, P_1 und $\alpha_2 = \frac{1}{2}$ gebildet.

Satz 9: *Q_k besteht aus den beiden abgeschlossenen Dreiecken $l_{k+1}l_kf_k$ und $h_{k+1}h_ki_k$ für $k \geqq 5$, bezw. $l_5a_4g_4$ und $h_5a_4c_4$ für $k = 4$ mit Ausnahme der nicht zu G_k gehörenden Ecken l_{k+1} und h_{k+1}. Die Seiten dieser Dreiecke werden von den Geraden $\alpha_1 + 2\alpha_2 = 1$, $\alpha_1 = 1/(k-1)$, $\alpha_2 = (k-1)/2k$ und $2\alpha_1 + \alpha_2 = 1$, $\alpha_1 = 1/(k-1)$, $\alpha_1 + \alpha_2 = (k-1)/k$ gebildet; für $k = 4$ tritt bei dem ersten dieser Dreiecke an die Stelle der Geraden $\alpha_1 = 1/(k-1)$ die Gerade $\alpha_1 + \alpha_2 = \frac{2}{3}$. Die Punktmenge Q_3 ist leer.*

Satz 10: *Es ist*

$$R_k = G_k - K_k - S_k - Q_k$$

unter Hinzunahme der in G_k auf H_k, P_2 und auf den Strecken $a_{k+1}b_k$, a_4d_3, a_kc_k, $l_{k+1}f_k$, l_5g_4, a_4g_4, $h_{k+1}i_k$, h_5c_4 und h_5i_5 gelegenen Randpunkten von K_k, S_k und Q_k mit Ausnahme der Punkte a_k und h_5.

3. **Folgerungen aus den Sätzen 2 und 3.** Obgleich die Formulierung der Sätze des vorigen Paragraphen etwas mühsam war, wurde sie hier doch aus folgenden Gründen genau durchgeführt:

Erstens macht ihr Ergebnis, die komplizierte Verteilung der Punktmengen K_k, S_k, Q_k und R_k die Richtigkeit der Vermutung (4) in hohem Maße wahrscheinlich. Denn aus den Sätzen 7–10 folgt, daß in jedem der in der Figur entstandenen Gebiete, mit Ausnahme von $a_3d_3u_3$, mindestens eine der den Rand bildenden Geraden die Eigenschaft hat, daß im Innern der Gebiete in hinreichender Nähe dieser Geraden der Unterschied der sich für γ ergebenden Abschätzung gegenüber der vermuteten Abschätzung $\gamma \geqq \alpha_1 + \alpha_2$ beliebig klein wird. So wird zum Beispiel nach Wahl einer beliebig kleinen positiven Größe ϵ in jedem der Gebiete R_k in hinreichender Nähe der Geraden $\alpha_1 = 1/k$, nämlich im Streifen

$$1/k < \alpha_1 < 1/k + \epsilon$$

der Unterschied der Abschätzungen kleiner als ϵ. Die analoge Eigenschaft haben die Geraden $\alpha_1 = \alpha_2$, $\alpha_1 + \alpha_2 = 1$, $\alpha_1 + 2\alpha_2 = 1$, $2\alpha_1 + \alpha_2 = 1$ und $\alpha_2 = 1/l$. Für kleine Werte von k liegen daher in R_k zahlreiche Parallelstreifen zur Achse $\alpha_2 = 0$, in denen der Unterschied der Abschätzung gegenüber (4) beliebig klein wird.

Zweitens erleichtert es die Anwendung des Satzes 5, wenn man bei zahlenmäßig gegebenen Werten von α_1 und α_2 sofort übersieht, welche der Abschätzungen das beste Resultat liefert. Insbesondere gilt dies, wenn der Satz 5 bei der Untersuchung der Summe von mehr als zwei Mengen successiv angewandt wird. Will man hier bei gegebenen Werten $\alpha_1, \alpha_2, \cdots, \alpha_n$ eine möglichst günstige Abschätzung erhalten, so muß, da die Addition der Mengen assoziativ und kommutativ ist, jede der möglichen Reihenfolgen, diese Mengen successiv zu addieren, untersucht werden. Denn ohne weiteres kann man nicht übersehen, das zeigen die Sätze 7–10, bei welcher Reihenfolge in der Ausführung der Additionen die Abschätzung auf Grund des Satzes 5 am günstigsten wird und welches Resultat man dann erhält. Für größere Werte von n hat man hier sehr viele Möglichkeiten zu betrachten.

Daher ist es, um die Anwendung der Sätze zu erleichtern, nicht überflüssig, durch Kombination der Sätze II, IV, 2 und 3 weitere Sätze abzuleiten, wenn auch diese Sätze im Falle $n = 2$ keine schärferen Resultate als Satz 5 liefern können. Insbesondere suchen wir geschlossene Abschätzungen für die Dichte der Summe von mehr als zwei Mengen zu erhalten, wie sie auch Satz 6 oder die direkte Anwendung von II auf n Mengen, also die Abschätzung

$$\gamma \geqq n\alpha_1 \tag{38}$$

ergeben.

Im Falle $n > 2$ kann die successive Anwendung von II schlechter oder besser als (38) sein. Zum Beispiel erhält man für drei Mengen A_1, A_2, A_3 mit den Dichten

$$0 < \alpha_1 = \alpha_2 = \alpha_3, \tag{39}$$

wenn $A_1 + A_2$ die Dichte β hat, einerseits aus (38)

$$\gamma \geqq 3\alpha_1, \tag{40}$$

andererseits $\beta \geqq 2\alpha_1 > \alpha_3$ und $\gamma \geqq 2\alpha_3$ durch successive Anwendung von II, also eine schlechtere Abschätzung als (40).

Gilt aber statt (39) $0 < \alpha_1 \leqq \frac{1}{2}\alpha_2 \leqq \frac{1}{2}\alpha_3$, so erhält man einerseits wieder (40), andererseits $\beta \geqq 2\alpha_1$ und $\gamma \geqq 4\alpha_1$ wegen $2\alpha_1 \leqq \alpha_3$. Dies ist besser als (40).

Ein einfaches Prinzip, durch Kombination von Sätzen weitere zu erhalten, ist das folgende: Kennt man zwei Abschätzungen

$$\gamma \geqq \phi_1(\alpha_1, \alpha_2, \cdots, \alpha_n),$$
$$\gamma \geqq \phi_2(\alpha_1, \alpha_2, \cdots, \alpha_n),$$

so gilt für jede Zahl a, die der Bedingung $0 < a < 1$ genügt,

$$a\gamma + (1 - a)\gamma = \gamma \geqq a\phi_1(\alpha_1, \alpha_2, \cdots, \alpha_n) + (1 - a)\phi_2(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

Die Dichten der Mengen A_1 und A_2 seien α_1 und α_2 und die von $A_1 + A_2$ sei $\gamma < 1$. Dann gelten zum Beispiel die folgenden Sätze:

Satz 11: *Es ist*

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{2}(1 + \alpha_2).$$

Beweis: Aus IV und Satz 3 folgt $\gamma \geqq \alpha_2/(1 - \alpha_1)$, $\gamma \geqq \alpha_2 + \alpha_1/(1 + \alpha_1)$. Wählt man $a = \frac{1}{2}(1 - \alpha_1)$, also $1 - a = \frac{1}{2}(1 + \alpha_1)$, so folgt

$$\gamma \geqq \tfrac{1}{2}\alpha_2 + \tfrac{1}{2}\alpha_2(1 + \alpha_1) + \tfrac{1}{2}\alpha_1 = \alpha_2 + \tfrac{1}{2}\alpha_1(1 + \alpha_2).$$

Satz 12: *Ist* $\alpha_1 \leqq \alpha_2$, *so gilt für jedes* a *mit* $0 < a < 1$

$$\gamma \geqq a\alpha_2 + \left(2 - \frac{3}{2}a\right)\alpha_1 + \frac{a}{2}\alpha_1\alpha_2.$$

Beweis: Nach Satz 11 und II ist

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{2} + \frac{\alpha_1\alpha_2}{2}, \qquad \gamma \geqq 2\alpha_1,$$

also

$$\gamma \geqq a\alpha_2 + \frac{a}{2}\alpha_1 + \frac{a}{2}\alpha_1\alpha_2 + 2(1 - a)\alpha_1 = a\alpha_2 + \left(2 - \frac{3}{2}a\right)\alpha_1 + \frac{a}{2}\alpha_1\alpha_2.$$

Speziell folgt für $a = \frac{4}{5}$

Satz 12a: *Es ist* $\gamma \geqq \frac{4}{5}(\alpha_1 + \alpha_2) + \frac{2}{5}\alpha_1\alpha_2$.

SATZ 13: *Für* $\alpha_1 \leqq \alpha_2$ *gilt* $\gamma \geqq \alpha_2 + 2\alpha_1^2$.

BEWEIS: Nach II und IV ist $\gamma \geqq 2\alpha_1$, $\gamma \geqq \alpha_2/(1-\alpha_1)$. Für $a = \alpha_1$ erhält man die Behauptung.

Diese Sätze kann man nun natürlich wieder kombinieren, usw.; hierauf soll nicht weiter eingegangen werden. Aus dem Satz 2 folgt ferner.

SATZ 14: *Die Dichten* α_1 *und* α_2 *der Mengen* A_1 *und* A_2 *mögen den Bedingungen* $\alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 \leqq 1$ *genügen; die ganze Zahl* $m \geqq 3$ *sei so bestimmt, daß*

$$\frac{m}{(m-1)^2} \geqq \alpha_1 \tag{41}$$

ist. Dann ist die Dichte γ *der Summe* $A_1 + A_2$

$$\gamma \geqq \alpha_2 + \frac{m-1}{m}\alpha_1.$$

BEWEIS: Für $\alpha_1 = 0$ ist die Behauptung trivial. Für $\alpha_1 > 0$ bestimme man wieder die ganze Zahl k so, daß

$$1/(k-1) \geqq \alpha_1 > 1/k \tag{42}$$

ist. Dann ist wegen (42) und (41)

$$1/k < \alpha_1 \leqq m/(m-1)^2 < 1/(m-2),$$

also $k > m-2$. Da k und m ganz sind, ist daher $k \geqq m-1$. Ist nun $k \geqq m$, so folgt aus Satz 2 und (42)

$$\gamma \geqq \alpha_2 + 1/k \geqq \alpha_2 + \alpha_1(k-1)/k \geqq \alpha_2 + \alpha_1(m-1)/m. \tag{43}$$

Ist aber $k = m-1$, so folgt aus Satz 2 und (41)

$$\gamma \geqq \alpha_2 + 1/k = \alpha_2 + 1/(m-1) \geqq \alpha_2 + \alpha_1(m-1)/m. \tag{44}$$

Aus (43) und (44) ergibt sich die Behauptung.

SATZ 15: *Es seien* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten*

$$\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n,$$

und es sei hierbei $n \geqq 2$; *die Dichte von* $A_2 + A_3 + \cdots + A_n$ *sei* β, *die von* $A_1 + A_2 + \cdots + A_n$ *sei* γ. *Ist dann* $\gamma < 1$, *so ist*

$$\gamma \geqq \beta + \alpha_1(n+1)/(n+2).$$

BEWEIS: Nach II ist $\alpha_1 < 1/n < (n+1)/n^2$ wegen $\gamma < 1$. Man muß nun zwei Fälle unterscheiden:

1) Es sei $\alpha_1 \leqq (n+2)/(n+1)^2$. Dann folgt aus Satz 14 für $m = n+2$ für die Dichte γ von $A_1 + (A_2 + A_3 + \cdots + A_n)$

$$\gamma \geqq \beta + \alpha_1(n+1)/(n+2).$$

2) Es sei

$$(n+2)/(n+1)^2 < \alpha_1 < 1/n. \tag{45}$$

Im Falle $n > 2$ kann die successive Anwendung von II schlechter oder besser als (38) sein. Zum Beispiel erhält man für drei Mengen A_1, A_2, A_3 mit den Dichten

$$0 < \alpha_1 = \alpha_2 = \alpha_3, \tag{39}$$

wenn $A_1 + A_2$ die Dichte β hat, einerseits aus (38)

$$\gamma \geqq 3\alpha_1, \tag{40}$$

andererseits $\beta \geqq 2\alpha_1 > \alpha_3$ und $\gamma \geqq 2\alpha_3$ durch successive Anwendung von II, also eine schlechtere Abschätzung als (40).

Gilt aber statt (39) $0 < \alpha_1 \leqq \frac{1}{2}\alpha_2 \leqq \frac{1}{2}\alpha_3$, so erhält man einerseits wieder (40), andererseits $\beta \geqq 2\alpha_1$ und $\gamma \geqq 4\alpha_1$ wegen $2\alpha_1 \leqq \alpha_3$. Dies ist besser als (40).

Ein einfaches Prinzip, durch Kombination von Sätzen weitere zu erhalten, ist das folgende: Kennt man zwei Abschätzungen

$$\gamma \geqq \phi_1(\alpha_1, \alpha_2, \cdots, \alpha_n),$$
$$\gamma \geqq \phi_2(\alpha_1, \alpha_2, \cdots, \alpha_n),$$

so gilt für jede Zahl a, die der Bedingung $0 < a < 1$ genügt,

$$a\gamma + (1-a)\gamma = \gamma \geqq a\phi_1(\alpha_1, \alpha_2, \cdots, \alpha_n) + (1-a)\phi_2(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

Die Dichten der Mengen A_1 und A_2 seien α_1 und α_2 und die von $A_1 + A_2$ sei $\gamma < 1$. Dann gelten zum Beispiel die folgenden Sätze:

Satz 11: *Es ist*

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{2}(1 + \alpha_2).$$

Beweis: Aus IV und Satz 3 folgt $\gamma \geqq \alpha_2/(1-\alpha_1)$, $\gamma \geqq \alpha_2 + \alpha_1/(1+\alpha_1)$. Wählt man $a = \frac{1}{2}(1-\alpha_1)$, also $1 - a = \frac{1}{2}(1+\alpha_1)$, so folgt

$$\gamma \geqq \tfrac{1}{2}\alpha_2 + \tfrac{1}{2}\alpha_2(1+\alpha_1) + \tfrac{1}{2}\alpha_1 = \alpha_2 + \tfrac{1}{2}\alpha_1(1+\alpha_2).$$

Satz 12: *Ist* $\alpha_1 \leqq \alpha_2$, *so gilt für jedes* a *mit* $0 < a < 1$

$$\gamma \geqq a\alpha_2 + \left(2 - \frac{3}{2}a\right)\alpha_1 + \frac{a}{2}\alpha_1\alpha_2.$$

Beweis: Nach Satz 11 und II ist

$$\gamma \geqq \alpha_2 + \frac{\alpha_1}{2} + \frac{\alpha_1\alpha_2}{2}, \qquad \gamma \geqq 2\alpha_1,$$

also

$$\gamma \geqq a\alpha_2 + \frac{a}{2}\alpha_1 + \frac{a}{2}\alpha_1\alpha_2 + 2(1-a)\alpha_1 = a\alpha_2 + \left(2 - \frac{3}{2}a\right)\alpha_1 + \frac{a}{2}\alpha_1\alpha_2.$$

Speziell folgt für $a = \frac{4}{5}$

Satz 12a: *Es ist* $\gamma \geqq \frac{4}{5}(\alpha_1 + \alpha_2) + \frac{2}{5}\alpha_1\alpha_2$.

SATZ 13: *Für* $\alpha_1 \leqq \alpha_2$ *gilt* $\gamma \geqq \alpha_2 + 2\alpha_1^2$.

BEWEIS: Nach II und IV ist $\gamma \geqq 2\alpha_1$, $\gamma \geqq \alpha_2/(1 - \alpha_1)$. Für $a = \alpha_1$ erhält man die Behauptung.

Diese Sätze kann man nun natürlich wieder kombinieren, usw.; hierauf soll nicht weiter eingegangen werden. Aus dem Satz 2 folgt ferner.

SATZ 14: *Die Dichten* α_1 *und* α_2 *der Mengen* A_1 *und* A_2 *mögen den Bedingungen* $\alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 \leqq 1$ *genügen; die ganze Zahl* $m \geqq 3$ *sei so bestimmt, daß*

$$\frac{m}{(m-1)^2} \geqq \alpha_1 \tag{41}$$

ist. Dann ist die Dichte γ *der Summe* $A_1 + A_2$

$$\gamma \geqq \alpha_2 + \frac{m-1}{m}\alpha_1.$$

BEWEIS: Für $\alpha_1 = 0$ ist die Behauptung trivial. Für $\alpha_1 > 0$ bestimme man wieder die ganze Zahl k so, daß

$$1/(k-1) \geqq \alpha_1 > 1/k \tag{42}$$

ist. Dann ist wegen (42) und (41)

$$1/k < \alpha_1 \leqq m/(m-1)^2 < 1/(m-2),$$

also $k > m - 2$. Da k und m ganz sind, ist daher $k \geqq m - 1$. Ist nun $k \geqq m$, so folgt aus Satz 2 und (42)

$$\gamma \geqq \alpha_2 + 1/k \geqq \alpha_2 + \alpha_1(k-1)/k \geqq \alpha_2 + \alpha_1(m-1)/m. \tag{43}$$

Ist aber $k = m - 1$, so folgt aus Satz 2 und (41)

$$\gamma \geqq \alpha_2 + 1/k = \alpha_2 + 1/(m-1) \geqq \alpha_2 + \alpha_1(m-1)/m. \tag{44}$$

Aus (43) und (44) ergibt sich die Behauptung.

SATZ 15: *Es seien* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten*

$$\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n,$$

und es sei hierbei $n \geqq 2$; *die Dichte von* $A_2 + A_3 + \cdots + A_n$ *sei* β, *die von* $A_1 + A_2 + \cdots + A_n$ *sei* γ. *Ist dann* $\gamma < 1$, *so ist*

$$\gamma \geqq \beta + \alpha_1(n+1)/(n+2).$$

BEWEIS: Nach II ist $\alpha_1 < 1/n < (n+1)/n^2$ wegen $\gamma < 1$. Man muß nun zwei Fälle unterscheiden:

1) Es sei $\alpha_1 \leqq (n+2)/(n+1)^2$. Dann folgt aus Satz 14 für $m = n + 2$ für die Dichte γ von $A_1 + (A_2 + A_3 + \cdots + A_n)$

$$\gamma \geqq \beta + \alpha_1(n+1)/(n+2).$$

2) Es sei

$$(n+2)/(n+1)^2 < \alpha_1 < 1/n. \tag{45}$$

Dann ist, da $A_1 + A_2 + \cdots + A_n$ wegen $\gamma < 1$ nicht die Menge aller positiven ganzen Zahlen ist, $\beta + \alpha_1 < 1$. Folglich ist

$$\beta + \alpha_1(n+1)/(n+2) = \beta + \alpha_1 - \alpha_1/(n+2) < 1 - \alpha_1/(n+2). \tag{46}$$

Nach (45) ist ferner $1 < \alpha_1(n+1)^2/(n+2)$. Daher folgt aus (46) und II

$$\beta + \frac{n+1}{n+2}\alpha_1 < \frac{(n+1)^2}{n+2}\alpha_1 - \frac{1}{n+2}\alpha_1 = \frac{n^2+2n}{n+2}\alpha_1 = n\alpha_1 \leqq \gamma.$$

Also gilt auch in diesem Falle $\gamma \geqq \beta + \alpha_1(n+1)/(n+2)$.

Für $n = 2$ liefert Satz 15 die folgende Verschärfung des Satzes 4:

SATZ 15a: *Ist* $\alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 < 1$, *so ist* $\gamma \geqq \alpha_2 + \frac{3}{4}\alpha_1$.

Durch successive Anwendung des Satzes 15 erhält man die folgende Verschärfung von VI:

SATZ 16: *Sind* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten* $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ *mit* $n \geqq 2$ *und hat* $A_1 + A_2 + \cdots + A_n$ *die Dichte* $\gamma < 1$, *so ist*

$$\gamma \geqq \alpha_n + \frac{3}{4}\alpha_{n-1} + \frac{4}{5}\alpha_{n-2} + \cdots + \frac{n+1}{n+2}\alpha_1.$$

BEWEIS: Für $n = 2$ ist die Behauptung mit Satz 15a identisch. Man nehme nun an, daß der Beweis für $n - 1$ Mengen schon erbracht sei. Ist dann β die Dichte der Summe der $n - 1$ Mengen $A_n, A_{n-1}, \cdots, A_2$, so ist also

$$\beta \geqq \alpha_n + \frac{3}{4}\alpha_{n-1} + \cdots + \frac{n}{n+1}\alpha_2.$$

Daher folgt aus Satz 15

$$\gamma \geqq \alpha_n + \frac{3}{4}\alpha_{n-1} + \cdots + \frac{n}{n+1}\alpha_2 + \frac{n+1}{n+2}\alpha_1.$$

SATZ 17: *Es seien* $A_1, A_2, \cdots, A_n$ *Mengen mit den Dichten*

$$\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n,$$

wo $n \geqq 2$ *ist; die Dichte von* $A_2 + A_3 + \cdots + A_n$ *sei* β, *die von*

$$A_1 + A_2 + \cdots + A_n$$

sei γ. *Ist dann* $\gamma < 1$, *so gilt für jede Zahl* a *mit* $0 < a < 1$

$$\gamma \geqq a\beta + \alpha_1\left\{n - \frac{n^2+n-1}{n+2}a\right\}.$$

BEWEIS: Es ist nach Satz 15 und II: $\gamma \geqq \beta + \alpha_1(n+1)/(n+2)$, $\gamma \geqq n\alpha_1$. Folglich ist

$$\gamma \geqq a\beta + \frac{n+1}{n+2}a\alpha_1 + (1-a)n\alpha_1 = a\beta + \alpha_1\left\{n - \frac{n^2+n-1}{n+2}a\right\}.$$

Satz 17 ist schärfer als VII, denn es ist

$$a\beta + \alpha_1\left\{n - \frac{n^2+n-1}{n+2}\,a\right\} \geqq a\beta + \alpha_1\{n - (n^2-n)^{\frac{1}{2}}\,a\},$$

weil wegen $n \geqq 2$

$$(n^2-n)(n+2)^2 = n^4 + 3n^3 - 4n > n^4 + 2n^3 - n^2 - 2n + 1 = (n^2+n-1)^2$$

ist.

Für $n = 2$ und $a = \frac{4}{5}$ liefert Satz 17 die folgende Verschärfung von VIIa:

SATZ 17a: *Ist* $\alpha_1 \leqq \alpha_2$ *und* $\alpha_1 + \alpha_2 < 1$, *so gilt*

$$\gamma \geqq \alpha_1 + \frac{4}{5}\alpha_2.$$

4. **Abschätzung der Dichte der Summe durch die Summe der Dichten.** Mit Hilfe des Satzes 15 läßt sich leicht der Satz VIII verbessern. Es gilt

SATZ 18: *Sind* $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ *die Dichten der Mengen* $A_1, A_2, \cdots, A_n$, *so gilt für die Dichte* γ *von* $A_1 + A_2 + \cdots + A_n$, *falls* $\gamma < 1$ *ist,*

$$\gamma \geqq \frac{8}{9}(\alpha_1+\alpha_2) > 0.8888\,(\alpha_1+\alpha_2) \qquad \text{für } n = 2,$$

$$\gamma \geqq \frac{120}{139}(\alpha_1+\alpha_2+\alpha_3) > 0.8633\,(\alpha_1+\alpha_2+\alpha_3) \qquad \text{für } n = 3,$$

$$\gamma \geqq \frac{2880}{3361}(\alpha_1+\alpha_2+\cdots+\alpha_n) > 0.8568\,(\alpha_1+\alpha_2+\cdots+\alpha_n) \qquad \text{für } n \geqq 4.$$

BEWEIS: 1) Nach Satz 15a ist für $n = 2$, $\gamma \geqq \alpha_2 + \frac{3}{4}\alpha_1$; andererseits ist nach II $\gamma \geqq 2\alpha_1$. Folglich ist

$$8\gamma + \gamma \geqq 8(\alpha_2 + \tfrac{3}{4}\alpha_1) + 2\alpha_1 = 8(\alpha_1+\alpha_2), \tag{47}$$

$$\gamma \geqq \tfrac{8}{9}(\alpha_1+\alpha_2).$$

2) Es sei nun $n = 3$. Die Dichte von $A_1 + A_2 + A_3$ sei $\gamma < 1$, die von $A_2 + A_3$ sei β. Dann ist nach (47) $\beta \geqq \frac{8}{9}(\alpha_2+\alpha_3)$, also nach Satz 15

$$\gamma \geqq \tfrac{8}{9}(\alpha_2+\alpha_3) + \tfrac{4}{5}\alpha_1.$$

Andererseits ist nach II $\gamma \geqq 3\alpha_1$, folglich

$$135\gamma + 4\gamma \geqq 135\{\tfrac{8}{9}(\alpha_2+\alpha_3) + \tfrac{4}{5}\alpha_1\} + 12\alpha_1 = 120(\alpha_1+\alpha_2+\alpha_3), \tag{48}$$

$$\gamma \geqq \frac{120}{139}(\alpha_1+\alpha_2+\alpha_3).$$

3) Es sei $n = 4$. Ist β die Dichte von $A_2 + A_3 + A_4$ und $\gamma < 1$ diejenige von $A_1 + A_2 + A_3 + A_4$, so ist nach Satz 15 und nach (48)

$$\gamma \geqq \beta + \frac{5}{6}\alpha_1 \geqq \frac{120}{139}(\alpha_2 + \alpha_3 + \alpha_4) + \frac{5}{6}\alpha_1.$$

Andererseits ist nach II $\gamma \geqq 4\alpha_1$, folglich

$$3336\gamma + 25\gamma \geqq 3336\left\{\frac{120}{139}(\alpha_2 + \alpha_3 + \alpha_4) + \frac{5}{6}\alpha_1\right\} + 100\,\alpha_1$$

$$= 2880(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4), \tag{49}$$

$$\gamma \geqq \frac{2880}{3361}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4).$$

4) Endlich sei $n > 4$. Aus Satz 16 und (49) folgt

$$\gamma \geqq \frac{2880}{3361}(\alpha_n + \alpha_{n-1} + \alpha_{n-2} + \alpha_{n-3}) + \frac{6}{7}\alpha_{n-4} + \frac{7}{8}\alpha_{n-5} + \cdots + \frac{n+1}{n+2}\alpha_1; \tag{50}$$

nun ist

$$\frac{2880}{3361} < \frac{6}{7}.$$

Daher folgt aus (50)

$$\gamma \geqq \frac{2880}{3361}(\alpha_1 + \alpha_2 + \cdots + \alpha_n).$$

BERLIN, GERMANY.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

NORMAL EXTENSIONS OF QUARTIC FIELDS WITH THE SYMMETRIC GROUP

BY D. M. DRIBIN[1]

(Received October 12, 1937)

1. Introduction

In an earlier[2] paper of the author, the Hilbert subgroup series in a normal algebraic number field N of degree 24 whose group is S_4, the symmetric group of degree four, were studied. The study of those subgroup series gave rise to a study of the characterization of those fields $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$, where B is normal of degree six with group S_3, the symmetric group of degree three, μ_1 and μ_2 are quantities in B, and N is the field described above. In the present paper necessary and sufficient conditions are given that $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ have the prescribed properties.

In §4 an interesting result is adduced concerning the extension of total-imaginary normal fields B with group S_3 to normal fields N with group S_4. In the final section the question of the realization of the various entries in the tables[3] of our earlier paper, is considered and it is shown that many of those cases are realizable.

It must be mentioned, finally, that in the entire paper, with the exception of §4, the reference field may be taken to be an arbitrary algebraic number field instead of R, the field of the rational numbers, as we do.

2. Groups of order 24 of a given type

If B is a normal algebraic number field with group S_3 and if $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ is normal with group S_4, then $N \mid B$ has as its Galois group the Vierer-group V. But not every such normal field $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ has S_4 as its group. We have, in fact, the following

THEOREM 1. *If $\mathfrak{G}$ is a group of order 24 which contains a normal subgroup isomorphic to V such that $\mathfrak{G}/V \cong S_3$, then $\mathfrak{G}$ is either (a) S_4, (b) $S_3 \times V$, (c) $\Phi = \{A, B, Q\}$, where*

$$A^4 = B^2 = Q^3 = 1, \qquad BAB = A^{-1}, \qquad A^{-1}QA = Q^{-1}, \qquad BQ = QB,$$

---

[1] National Research Fellow.

[2] *Quartic fields with the symmetric group*, *Annals of Math.* (2), 38 (1937) pp. 739–749. We shall refer to this paper as Q.

[3] A number of cases occurring under Table C seem to have been inadvertently omitted; they are briefly given in §5.

or (*d*) $\Psi = \{A, B, Q\}$, *where*

$$A^4 = B^2 = Q^3 = 1, \quad AB = BA, \quad A^{-1}QA = Q^{-1}, \quad BQ = QB.$$

The proof depends upon the known list of groups of order 24 and can be given quite easily with its use.

There are twelve[4] *non-abelian*[5] groups of order 24, given in terms of their generators as follows:

(I) $A^4 = B^4 = Q^3 = 1, B^2 = A^2, B^{-1}AB = A^{-1}, AQ = QA, BQ = QB;$

(II) $A^8 = Q^3 = 1, A^{-1}QA = Q^{-1};$

(III) $A^4 = B^4 = Q^3 = 1, B^2 = A^2, B^{-1}AB = A^{-1}, AQ = QA, B^{-1}QB = Q^{-1};$

(IV) $A^4 = B^4 = Q^3 = 1, B^2 = A^2, B^{-1}AB = A^{-1}, Q^{-1}AQ = B, Q^{-1}BQ = AB;$

(V) $A^4 = B^2 = Q^3 = 1, AB = BA, AQ = QA, BQB = Q^{-1};$

(VI) $A^4 = B^2 = Q^3 = 1, BAB = A^{-1}, AQ = QA, BQB = Q^{-1};$

(VII) $A^4 = B^2 = Q^3 = 1, BAB = A^{-1}, AQ = QA, BQ = QB;$

(VIII) $A^2 = B^2 = C^2 = Q^3 = 1, AB = BA, AC = CA, BC = CB, AQ = QA, Q^{-1}BQ = C, Q^{-1}CQ = BC;$

(IX) $A^4 = B^2 = Q^3 = 1, BAB = A^{-1}, A^{-1}QA = Q^{-1}, BQ = QB;$

(X) $A^2 = B^2 = C^2 = Q^3 = 1, AB = BA, AC = CA, BC = CB, AQ = QA, BQ = QB, CQC = Q^{-1};$

(XI) $A^4 = B^2 = Q^3 = 1, BAB = A^{-1}, Q^{-1}A^2Q = B, Q^{-1}BQ = A^2B, A^{-1}QA = Q^2A^2B;$

(XII) $A^4 = B^2 = Q^3 = 1, AB = BA, A^{-1}QA = Q^{-1}, BQ = QB.$

In cases (I) to (IV) there is only one element of order two, namely, A^2, A^4, A^2, and A^2, respectively. Hence none of these groups can have V as a subgroup.

In case (V) there are three subgroups isomorphic to V, namely, the subgroups $(1, A^2, B, A^2B)$, $(1, A^2, BQ, A^2BQ)$, and $(1, A^2, BQ^2, A^2BQ^2)$. But none of these is invariant in the larger group. In case (VI) the group contains no invariant subgroup of type V.

In case (VII) there are two invariant subgroups of type V, namely, $\mathfrak{H} = (1, A^2, B, A^2B)$ and $\mathfrak{H}' = (1, A^2, AB, A^3B)$. Decomposing $\mathfrak{G}$ relative to $\mathfrak{H}$ we find the elements of the factor group to be $\mathfrak{H}$, $\mathfrak{H}A$, $\mathfrak{H}Q$, $\mathfrak{H}Q^2$, $\mathfrak{H}AQ$, $\mathfrak{H}AQ^2$, with $\mathfrak{H} = (\mathfrak{H}A)^2 = (\mathfrak{H}Q)^3 = (\mathfrak{H}Q^2)^3 = (\mathfrak{H}AQ)^3 = (\mathfrak{H}AQ^2)^3$. But this factor group is not isomorphic to S_3. Similarly, the factor group $\mathfrak{G}/\mathfrak{H}'$ is not isomorphic to S_3. Hence case (VII) does not have the desired properties.

The only invariant subgroup of type V in case (VIII) is $(1, B, C, BC)$ but,

---

[4] A. E. Western, *Groups of order p^3q*, *Proc. Lond. Math. Soc.* (1) (1899) 30, pp. 209–263.

[5] There are three abelian groups of order 24 but they are excluded since S_3, and hence $\mathfrak{G}$, is non-abelian.

as before, we find that the factor group has only one element of order two and cannot, then, be an isomorph of S_3. Case (VIII), then, cannot yield the desired properties.

The remaining four cases, however, all have the desired properties. Case (IX) is Φ of the theorem, (X) is $S_3 \times V$, (XI) is S_4, and (XII) is Ψ. Hence we have the statement in Theorem 1.

All the groups given in Theorem 1, with the sole exception of S_4, possess an invariant subgroup of order two: for $S_3 \times V$, it is the subgroup $(1, A)$; for Φ and Ψ, it is $(1, A^2)$. Since the corresponding normal field of degree 24 would then possess a normal subfield of degree 12, we have the

THEOREM 2. *Let $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ be a normal field of degree 24 over R, where B is a normal algebraic number field with group S_3. Then $N \mid R$ has S_4 as Galois group if and only if no one of the three fields $B(\sqrt{\mu_1})$, $B(\sqrt{\mu_2})$, $B(\sqrt{\mu_1\mu_2})$ is normal over R.*

3. Class field characterization of N/R

Let B, as before, be a normal algebraic number field with group S_3, and let $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$, where $\sqrt{\mu_1}$ and $\sqrt{\mu_2}$ are independent quantities, so that $[N:B] = 4$. Since $N \mid B$ is abelian, N is a class field over B for a certain ideal group $\mathfrak{H}$ in B. By the class field theory it follows that $N \mid R$ is normal if and only if $\mathfrak{H}^S = \mathfrak{H}$ for every automorphism S of S_3. This condition, however, is merely necessary and is not sufficient to determine the group of $N \mid R$; these further conditions we now seek. We shall denote by σ and τ the generating elements of S_3: $\sigma^2 = \tau^3 = 1$, $\sigma\tau = \tau^2\sigma$.

If $\mathfrak{A}$ represents the totality of the ideals in B that are prime to the conductor of $\mathfrak{H}$, then by the isomorphism theorem of the class field theory, $\mathfrak{A} = \mathfrak{H} + \mathfrak{C}_1 + \mathfrak{C}_2 + \mathfrak{C}_1\mathfrak{C}_2$. We shall prove the

THEOREM 3. *A necessary and sufficient condition that $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ be normal over R with group S_4 is that*

$$(1) \qquad \mathfrak{C}_1^{\bar\sigma} = \mathfrak{C}_1\mathfrak{C}_2, \qquad \mathfrak{C}_1^{\bar\tau} = \mathfrak{C}_2, \qquad \mathfrak{C}_2^{\bar\sigma} = \mathfrak{C}_2, \qquad \mathfrak{C}_2^{\bar\tau} = \mathfrak{C}_1\mathfrak{C}_2$$

for a properly chosen pair of generating automorphisms $\bar\sigma$ and $\bar\tau$ of the Galois group of B.[5a]

Since, if $\mathfrak{C}$ is a class of $\mathfrak{A}/\mathfrak{H}$, the ideal group $\mathfrak{H} + \mathfrak{C}$ determines a quadratic field $B(\sqrt{\mu})$ over B, so that the Kummer theory yields by (1) the equivalent

THEOREM 4. *A necessary and sufficient condition that $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ be normal over R with group S_4 is that*

$$(2) \qquad \mu_1^{\bar\sigma} \sim \mu_1\mu_2, \qquad \mu_1^{\bar\tau} \sim \mu_2, \qquad \mu_2^{\bar\sigma} \sim \mu_2, \qquad \mu_2^{\bar\tau} \sim \mu_1\mu_2$$

for a properly chosen pair of generating automorphisms $\bar\sigma$ and $\bar\tau$ of the Galois group of B. The equivalence relation "$\sim$" is determined by the class of all squares of elements of B; that is, $\alpha_1 \sim \alpha_2$ means $\alpha_1 = \alpha^2\alpha_2$ where α is in B.

[5a] It follows immediately by (1) that $\mathfrak{H}$ is necessarily invariant under the automorphisms of B/R. For this reason the invariance of $\mathfrak{H}$ is not mentioned in Theorems 4 and 6.

We proceed to the proof of Theorem 3. We note again that by virtue of Theorem 2, $\mathfrak{H} + \mathfrak{C}$ is not invariant under S_3, where $\mathfrak{C}$ is any class $\neq \mathfrak{H}$ in $\mathfrak{A}/\mathfrak{H}$.

(a) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1$, $\mathfrak{C}_1^\tau = \mathfrak{C}_2$. Hence $\mathfrak{C}_2^\sigma \neq \mathfrak{C}_1$, $\mathfrak{C}_2^\tau \neq \mathfrak{C}_2$ for otherwise, $\mathfrak{C}_1 = \mathfrak{C}_2$. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, then $\mathfrak{C}_1^{\tau^3} = \mathfrak{C}_2$ and $\mathfrak{C}_1 = \mathfrak{C}_2$. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, then $\mathfrak{C}_1^{\sigma\tau} = \mathfrak{C}_2$, $\mathfrak{C}_1^{\tau^2\sigma} = \mathfrak{C}_1\mathfrak{C}_2$ and $\mathfrak{C}_2 = \mathfrak{C}_1\mathfrak{C}_2$. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, then $\mathfrak{C}_1^{\tau^2} = \mathfrak{C}_1$, whence $\mathfrak{C}_1 = \mathfrak{C}_2$. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, then we get (1) if we write $\bar\sigma = \sigma\tau^2$, $\bar\tau = \tau$; for, then,

$$\mathfrak{C}_1^{\bar\sigma} = \mathfrak{C}_1\mathfrak{C}_2, \qquad \mathfrak{C}_1^{\bar\tau} = \mathfrak{C}_2, \qquad \mathfrak{C}_2^{\bar\sigma} = \mathfrak{C}_2, \qquad \mathfrak{C}_2^{\bar\tau} = \mathfrak{C}_1\mathfrak{C}_2.$$

(b) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1$, $\mathfrak{C}_1^\tau = \mathfrak{C}_1\mathfrak{C}_2$. Hence $\mathfrak{C}_2^\sigma \neq \mathfrak{C}_1$, $\mathfrak{C}_2^\tau \neq \mathfrak{C}_1\mathfrak{C}_2$. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, then $\mathfrak{C}_1^{\sigma\tau} = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_1^{\tau^2\sigma} = \mathfrak{C}_2$, which is impossible. If $\mathfrak{C}_2^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, we get (1) with $\bar\sigma = \sigma\tau$, $\bar\tau = \tau^2$. Finally, if $\mathfrak{C}_2^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_2$, we get $\mathfrak{C}_1 = \mathfrak{C}_1^{\tau^3} = (\mathfrak{C}_1\mathfrak{C}_2)^{\tau^2} = \mathfrak{C}_1^\tau\mathfrak{C}_2^\tau\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, a contradiction.

(c) $\mathfrak{C}_1^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_1^\tau = \mathfrak{C}_1$. Hence $\mathfrak{C}_2^\sigma = \mathfrak{C}_1$, $\mathfrak{C}_2^\tau \neq \mathfrak{C}_1$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_2$, then $(\mathfrak{C}_1\mathfrak{C}_2)^\sigma = \mathfrak{C}_1\mathfrak{C}_2 = (\mathfrak{C}_1\mathfrak{C}_2)^\tau$, contrary to the structure of $N \mid R$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, then $\mathfrak{C}_2 = \mathfrak{C}_2^{\tau^3} = (\mathfrak{C}_1\mathfrak{C}_2)^{\tau^2} = \mathfrak{C}_1^\tau(\mathfrak{C}_1\mathfrak{C}_2)^\tau = \mathfrak{C}_1\mathfrak{C}_2$, a contradiction.

(d) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1^\tau = \mathfrak{C}_2$. Hence $\mathfrak{C}_2^\sigma = \mathfrak{C}_1$, $\mathfrak{C}_2^\tau \neq \mathfrak{C}_2$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, then $(\mathfrak{C}_1\mathfrak{C}_2)^\sigma = (\mathfrak{C}_1\mathfrak{C}_2)^\tau = \mathfrak{C}_1\mathfrak{C}_2$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, then we get (1) with $\bar\sigma = \sigma\tau$, $\bar\tau = \tau$.

(e) $\mathfrak{C}_1^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_1^\tau = \mathfrak{C}_1\mathfrak{C}_2$. Hence $\mathfrak{C}_2^\sigma = \mathfrak{C}_1$, $\mathfrak{C}_2^\tau \neq \mathfrak{C}_1\mathfrak{C}_2$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, we get (1) with $\bar\sigma = \sigma\tau^2$, $\bar\tau = \tau^2$. But if $\mathfrak{C}_2^\tau = \mathfrak{C}_2$, we get $\mathfrak{C}_1 = \mathfrak{C}_1^{\tau^3} = (\mathfrak{C}_1\mathfrak{C}_2)^{\tau^2} = (\mathfrak{C}_1\mathfrak{C}_2)^\tau\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, a contradiction.

(f) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_1^\tau = \mathfrak{C}_1$. We have $\mathfrak{C}_1 = \mathfrak{C}_1^{\sigma^2} = \mathfrak{C}_1\mathfrak{C}_2\mathfrak{C}_2^\sigma$ whence $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$; therefore $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$. But then $\mathfrak{C}_2 = \mathfrak{C}_2^{\tau^3} = (\mathfrak{C}_1\mathfrak{C}_2)^{\tau^2} = (\mathfrak{C}_1\mathfrak{C}_2)^\tau\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$, a contradiction.

(g) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1\mathfrak{C}_2$, $\mathfrak{C}_1^\tau = \mathfrak{C}_2$. As above, $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$ and $\mathfrak{C}_2^\tau \neq \mathfrak{C}_2$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1$, then $\mathfrak{C}_1 = \mathfrak{C}_1^{\tau^3} = \mathfrak{C}_2^{\tau^2} = \mathfrak{C}_1^\tau = \mathfrak{C}_2$. If $\mathfrak{C}_2^\tau = \mathfrak{C}_1\mathfrak{C}_2$ we get (1) with $\bar\sigma = \sigma$, $\bar\tau = \tau$.

(h) $\mathfrak{C}_1^\sigma = \mathfrak{C}_1^\tau = \mathfrak{C}_1\mathfrak{C}_2$. As above, $\mathfrak{C}_2^\sigma = \mathfrak{C}_2$, $\mathfrak{C}_2^\tau = \mathfrak{C}_1$. We get (1) with $\bar\sigma = \sigma$, $\bar\tau = \tau^2$.

Hence Theorem 2 (and therefore Theorem 3) is completely proved.

4. The case B (total-)imaginary field over R

We shall first prove the following sufficient conditions for Theorem 3:

THEOREM 5. *The conditions* (2) *of Theorem* 4 *will be fulfilled if there exists a quantity* $\gamma \neq 0$ *in* B *which is not a square in* B *but which satisfies the conditions*

$$\gamma\gamma^\sigma = \xi^2, \qquad \gamma\gamma^\tau\gamma^{\tau^2} = \eta^2, \tag{3}$$

where ξ *and* η *are quantities of* B.

We take $\mu_2 = \gamma$. Then $\mu_2^\sigma = \xi^2/\mu_2 = (\xi/\mu_2)^2\mu_2 \sim \mu_2$, which yields (2_3).

We write $\mu_1 = \mu_2^{\tau^2}$ so that $\mu_1^\tau = \mu_2$ and (2_2) is satisfied. Then

$$\mu_1^\sigma = \mu_2^{\tau^2\sigma} = \mu_2^{\sigma\tau} \sim 1/\mu_2^\tau = (1/\mu_2\mu_2^\tau\mu_2^{\tau^2})(\mu_2\mu_2^{\tau^2}) = \frac{1}{\eta^2}\mu_1\mu_2 \sim \mu_1\mu_2,$$

which yields (2_1).

Finally, $\mu_2^\tau = \eta^2/\mu_2\mu_2^{\tau^2} \sim \mu_2\mu_2^{\tau^2} = \mu_1\mu_2$ and condition (2_4) is satisfied. Hence if $\mu_1 \neq \alpha^2\mu_2$ with α in B, Theorem 4 will certainly hold. But if $\mu_2 \sim \mu_2^{\tau^2}$, then $\mu_2 \sim \mu_2^\tau$, so that $\mu_2 \sim \mu_2\mu_2^\tau\mu_2^{\tau^2} = \gamma\gamma^\tau\gamma^{\tau^2} \sim 1$ whereas $\mu_2 = \gamma$ is not a square.

In the remainder of this section we shall show that if the (unique) quadratic field $R(\sqrt{d})$ that is contained in B is imaginary, we can extend B to a normal field $N \mid R$ of degree 24 over R and whose absolute Galois group is S_4 and that, in fact, we can choose the quantities μ_1 and μ_2 that are involved in the adjunction to be units in B.

Since B is now assumed to be a total-imaginary field of degree 6, there exist, by the Dirichlet theorem on the units of an algebraic number field, two fundamental units ϵ_1 and ϵ_2 in B such that every unit ϵ in B can be expressed uniquely in the form

$$\epsilon = \zeta^e \epsilon_1^{e_1} \epsilon_2^{e_2}$$

where ζ is a primitive w-th root of unity contained in B (whence $w \leqq 7$) and e, e_1, e_2 are integers.

It is easily seen that $\zeta^e \sim \pm 1$ except when $w = 4$ and that in that event $\zeta^e \sim i^a (a = 0 \text{ or } 1)$, where $i = \sqrt{(-1)}$. For, if $w = 2^a m$, m odd, we have, if $\alpha = 0$, $\zeta = \zeta^{m+1} = (\zeta^{\frac{1}{2}(m+1)})^2 \sim 1$; if $\alpha = 1$, $\zeta^2 = \zeta^{2m+2}$, whence $\zeta = \pm(\zeta^{\frac{1}{2}(m+1)})^2 \sim \pm 1$; if $\alpha = 2$, $\zeta^4 = \zeta^{4m+4}$, whence $\zeta = i^b(\zeta^{\frac{1}{2}(m+1)})^2 \sim 1$ or i according as b is even or odd.

As a convenience we shall use in the following a second symbol of equivalence; we shall write $\epsilon \approx \eta$ if $\epsilon \sim i^a \eta$. We shall now show that γ of Theorem 5 can be chosen $\approx \epsilon_1$, ϵ_2, or $\epsilon_1\epsilon_2$.

Let $\epsilon_1^r \approx \epsilon_1^{b_1}\epsilon_2^{b_2}, \epsilon_2^r \approx \epsilon_1^{b_1'}\epsilon_2^{b_2'}$, where the b_i, b_i' are taken to be reduced modulo 2. We list in the adjoining table the possible combinations of b_1, b_2, b_1', b_2' modulo 2, denoting by an asterisk (*) an arbitrary choice of 0 or 1.

	b_1	b_2	b_1'	b_2'
1	0	0	*	*
2	*	*	0	0
3	0	1	0	1
4	0	1	1	0
5	0	1	1	1
6	1	0	0	1
7	1	0	1	0
8	1	0	1	1
9	1	1	0	1
10	1	1	1	0
11	1	1	1	1

Combination (1) cannot occur, for if $\epsilon_1^\tau \approx 1$, then $\epsilon_1 \approx 1$, an impossibility, since ϵ_1 is a fundamental unit; similarly, (2) cannot occur. (3) yields $\epsilon_1^\tau \approx \epsilon_2 \approx \epsilon_2^\tau$, whence $\epsilon_1 \approx \epsilon_2$, which is impossible, since ϵ_1 and ϵ_2 are *independent* fundamental units in B.

(4) yields $\epsilon_1^\tau \approx \epsilon_2$, $\epsilon_2^\tau \approx \epsilon_1$, whence $\epsilon_1^{\tau^2} \approx \epsilon_1$ and $\epsilon_1 = \epsilon_1^{\tau^3} \approx \epsilon_1^\tau \approx \epsilon_2$, an impossibility.

In case (5) we must revert to the earlier equivalence relation. We can now choose γ satisfying $(3)_2$ of Theorem 4. If $\epsilon_1^\tau \sim \pm\epsilon_2$, $\epsilon_2^\tau \sim \epsilon_1\epsilon_2$, we take $\gamma = \epsilon_1$, ϵ_2, or $\epsilon_1\epsilon_2$. If $\epsilon_1^\tau \sim \epsilon_2$, $\epsilon_2^\tau \sim -\epsilon_1\epsilon_2$, we take $\gamma = -\epsilon_1$, $-\epsilon_2$, or $\epsilon_1\epsilon_2$. If $\epsilon_1^\tau \sim -\epsilon_2$, $\epsilon_2^\tau \sim \epsilon_1\epsilon_2$, we take $\gamma = \epsilon_1$, $-\epsilon_2$, or $-\epsilon_1\epsilon_2$. If $\epsilon_1^\tau \sim \epsilon_2$, $\epsilon_2^\tau \sim i\epsilon_1\epsilon_2$, we take $\gamma = i\epsilon_1$, $i\epsilon_2$, or $-\epsilon_1\epsilon_2$. If $\epsilon_1^\tau \sim i\epsilon_2$, $\epsilon_2^\tau \sim \epsilon_1\epsilon_2$, we take $\gamma = \epsilon_1$, $i\epsilon_2$, or $i\epsilon_1\epsilon_2$. If $\epsilon_1^\tau \sim i\epsilon_2$, $\epsilon_2^\tau \sim i\epsilon_1\epsilon_2$, we take $\gamma = i\epsilon_1$, ϵ_2, or $i\epsilon_1\epsilon_2$.

In case (6) we get

$$\epsilon_1 = \epsilon_1^{\tau^3} = \zeta^\alpha \epsilon_1^{\beta_1(\beta_1^2+\beta_1'\beta_2)+\beta_1'\beta_2(\beta_1+\beta_2')} \epsilon_2^{\beta_2(\beta_1^2+\beta_1'\beta_2)+\beta_2\beta_2'(\beta_1+\beta_2')},$$

where $\beta_i \equiv b_i$, $\beta_i' \equiv b_i'$ (mod 2), and we must have

$$\begin{aligned} \beta_1(\beta_1^2 + \beta_1'\beta_2) + \beta_1'\beta_2(\beta_1 + \beta_2') &= 1 \\ \beta_2(\beta_1^2 + \beta_1'\beta_2) + \beta_2\beta_2'(\beta_1 + \beta_2') &= 0. \end{aligned} \tag{4}$$

If $\beta_2 \neq 0$ these equations yield

$$\beta_1 + \beta_2' = 1/(\beta_1'\beta_2 - \beta_1\beta_2')$$

and since the β_i and β_i' are integers, $\beta_1 + \beta_2' = \pm 1$, whereas we are assuming $\beta_1\beta_2' \equiv 1$ (mod 2).[6]

Hence $\beta_2 = 0$ and equations (4) yield $\beta_1 = 1$. Therefore $\epsilon_1^\tau = \zeta^\beta\epsilon_1$ and $(\epsilon_1^w)^\tau = \epsilon_1^w$. But then ϵ_1^w is a unit in $R(\sqrt{d})$ and hence must be a root of unity since we have assumed $d < 0$. This, however, is impossible, from the nature of ϵ_1 and so case (6) cannot exist.

In case (7) $\epsilon_1^\tau \approx \epsilon_1$, $\epsilon_2^\tau \approx \epsilon_1$, whence $\epsilon_1^\tau \approx \epsilon_2^\tau$ and $\epsilon_1 \approx \epsilon_2$.

In (8) $\epsilon_1^\tau \approx \epsilon_1$, $\epsilon_2^\tau \approx \epsilon_2$. We find $\epsilon_2 = \epsilon_2^{\tau^3} \approx \epsilon_1\epsilon_2$ whence $\epsilon_2 \approx 1$; hence this case cannot exist.

In (10) we proceed as in case (5). We need merely note that if ϵ_1 and ϵ_2 be interchanged, we get case (5).

We now seek to satisfy $(3)_1$ by a γ just determined that satisfies $(3)_2$. We write $\epsilon_1^\sigma \approx \epsilon_1^{a_1}\epsilon_2^{a_2}$, $\epsilon_2^\sigma \approx \epsilon_1^{a_1'}\epsilon_2^{a_2'}$, where a_i, a_i' are reduced modulo 2. We consider *seriatim* the cases in the table where we now read a for b.

Cases (1) and (2) are again impossible. Case (3) implies $\epsilon_1^\sigma \approx \epsilon_2$, $\epsilon_2^\sigma \approx \epsilon_2$, whence $\epsilon_1 \approx \epsilon_2$, an impossibility.[7]

In case (4) we have $\epsilon_1^\sigma \sim \epsilon_2$, $-\epsilon_2$, or $i\epsilon_2$, whence, since $\sigma^2 = 1$, $\epsilon_2^\sigma \sim \epsilon_1$, $-\epsilon_1$, or $i\epsilon_1$. Hence $(\epsilon_1\epsilon_2)(\epsilon_1\epsilon_2)^\sigma \sim 1$ in all cases and we may take $\gamma = \epsilon_1\epsilon_2$. Since

[6] This remark disposes of the cases (9) and (11) since then $\beta_2 \neq 0$.

[7] We note that in the event i is in B, $i^\sigma = -i \sim i$.

multiplication by -1 or i does not affect the validity of $(3)_1$, we may take γ to satisfy $(3)_2$ as well.

In case (5) we have $\epsilon_1^\sigma \approx \epsilon_2$, $\epsilon_2^\sigma \approx \epsilon_1\epsilon_2$. Therefore, $\epsilon_1 = \epsilon_1^{\sigma^2} \approx \epsilon_1\epsilon_2$ whence $\epsilon_2 \approx 1$, an impossibility.

In case (6) we have $\epsilon_1^\sigma \approx \epsilon_1$, $\epsilon_2^\sigma \approx \epsilon_2$. If $\epsilon_1^\sigma \sim \epsilon_1$, then $\epsilon_1\epsilon_1^\sigma \sim 1$ and we take $\gamma = \epsilon_1$; if $\epsilon_2^\sigma \sim \epsilon_2$, then we take $\gamma = \epsilon_2$; if $\epsilon_1^\sigma \sim -\epsilon_1$, $\epsilon_2^\sigma \sim -\epsilon_2$, we take $\gamma = \epsilon_1\epsilon_2$; if $\epsilon_1^\sigma \sim i\epsilon_1$, $\epsilon_2^\sigma \sim i\epsilon_2$, then $(\epsilon_1\epsilon_2)(\epsilon_1\epsilon_2)^\sigma \sim 1$, and we take $\gamma = \epsilon_1\epsilon_2$. As was remarked above, multiplication of γ by -1 or i does not affect the validity of $(3)_1$.

Case (7) is impossible since $\epsilon_1^\sigma \approx \epsilon_1$, $\epsilon_2^\sigma \approx \epsilon_1$, implies $\epsilon_1 \approx \epsilon_2$.

In case (8) $\epsilon_1^\sigma \approx \epsilon_1$, $\epsilon_2^\sigma \approx \epsilon_1\epsilon_2$. If $\epsilon_1^\sigma \sim \epsilon_1$, we take $\gamma = \epsilon_1$; if $\epsilon_1^\sigma \sim -\epsilon_1$, then $\epsilon_2^\sigma \sim -\epsilon_1\epsilon_2$ and $(\epsilon_1\epsilon_2)^{\sigma^2} \sim -\epsilon_1\epsilon_2$, which is false; if $\epsilon_1^\sigma \sim i\epsilon_1$, $\epsilon_2^\sigma \sim \epsilon_1\epsilon_2$, then $(\epsilon_1\epsilon_2)^{\sigma^2} \sim i\epsilon_1\epsilon_2$, which is also false; finally, if $\epsilon_1^\sigma \sim i\epsilon_1$, $\epsilon_2^\sigma \sim i\epsilon_1\epsilon_2$, then $(\epsilon_1\epsilon_2)^{\sigma^2} \sim i\epsilon_1\epsilon_2$——again false.

Case (9) can be treated as case (8) with ϵ_1 and ϵ_2 interchanged.

In case (10) $\epsilon_1^\sigma \approx \epsilon_1\epsilon_2$, $\epsilon_2^\sigma \approx \epsilon_1$, whence $\epsilon_2 = \epsilon_2^{\sigma^2} \approx \epsilon_1\epsilon_2$ and $\epsilon_1 \approx 1$, a contradiction.

Finally, (11) is impossible since it leads to $\epsilon_1^\sigma \approx \epsilon_2$ and $\epsilon_1 \approx \epsilon_2$.

We have thus proved the

THEOREM 6. *If B is a normal sextic field containing an imaginary quadratic field $R(\sqrt{d})$, there exist units μ_1 and μ_2 in B such that $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ is a normal field of degree 24 over R whose group is isomorphic to S_4.*

5. Application to Hilbert subgroup series of $N \mid R$

We shall prove very simply the

THEOREM 7. *There exist normal fields B and $N \supset B$ with groups S_3 and S_4, respectively, and infinitely many prime ideals $\mathfrak{p}$ in B such that any Hilbert subgroup series for N in Table*[8] *A is realizable.*

If $R(\sqrt{d})$ is an imaginary quadratic field there exist[9] infinitely many normal fields of degree six over R and with group S_3; let B be such a field. We now choose $N = B(\sqrt{\mu_1}, \sqrt{\mu_2})$ as in Theorem 5. We denote by $\mathfrak{H}_1$ and $\mathfrak{H}_2$ the ideal groups of index two in B which correspond to $B(\sqrt{\mu_1})$ and $B(\sqrt{\mu_2})$ as class fields over B; we also denote by $\mathfrak{k}$ the ideal group of index three in $R(\sqrt{d})$ corresponding to the cyclic field $B \mid R(\sqrt{d})$. (We shall understand $B(\sqrt{\mu_1})$ to be the field Λ_1 in the tables.)

We have the following conditions, which follow immediately from the tables of Q.:

$(A, 1)$ *is realizable for $\mathfrak{p}$ in $\mathfrak{H}_2$ and $\mathfrak{k}$, with* $\left(\frac{d}{p}\right) = +1$,

$(A, 2)$ *is realizable for $\mathfrak{p}$ not in $\mathfrak{H}_2$ but in $\mathfrak{k}$, with* $\left(\frac{d}{p}\right) = -1$,

---

[8] This refers to the tables in Q.

[9] Hasse, *Arithmetische Theorie der kubischen Zahlkörper auf klassenkörpertheoretischer Grundlage*, *Math. Zeitschr.* 31 (1930), pp. 565–582.

$(A, 3)$ *is realizable for* $\mathfrak{p}$ *in* $\mathfrak{H}_2$ *but not in* $\mathfrak{L}$,

$(A, 4)$ *is realizable for* $\mathfrak{p}$ *not in* $\mathfrak{H}_2$ *but in* $\mathfrak{L}$, *with* $\left(\frac{d}{p}\right) = +1$,

$(A, 5)$ *is realizable for* $\mathfrak{p}$ *in* $\mathfrak{H}_2$ *and* $\mathfrak{L}$, *with* $\left(\frac{d}{p}\right) = -1$, where p, the rational prime divisible by p in $\mathfrak{H}_1$ is odd and prime to the discriminant of $B \mid R$. There exist infinitely many prime ideals satisfying any one of these conditions, as is well known from the theory of class fields.

In Table B we can show in a similar manner the existence of $(B, 1)$ and $(B, 6)$; the existence of $(B, 5)$ and $(B, 10)$ can be shown by the exhibition of concrete cases of N, since our methods do not yield the existence of more than one of these cases.

For $(B, 5)$ we take $p = 7$ and $f(x) = x^4 + 90x^3 + 188x^2 + 85x + 105$ where $f(x) = 0$ defines a quartic field K whose group can readily be seen to be[10] S_4. $f(x) \equiv x(x + 1)(x - 1)^2 \pmod 7$, so that since 7 can be easily shown to be prime to the index of the field by a criterion[11] of Dedekind, it follows by another[11] theorem of Dedekind that in K, $7 = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3^2$, where the $\mathfrak{p}_i$ are prime ideals in K of degree 1. Such a case can occur only for $(B, 5)$ and hence this case is realizable.

The case $(B, 10)$ is treated similarly but with $p = 5$ and

$$f(x) = x^4 - 22x^2 - 5x - 15.$$

As above, it follows, since $f(x) \equiv x^2(x^2 - 2) \pmod 5$, where $x^2 - 2$ is irreducible modulo 5, and since 5 is prime to the index of K (again by the criterion of Dedekind) that $5 = \mathfrak{p}_1^2\mathfrak{p}_2^2$ in K. This can occur only for $(B, 10)$ and hence this case is realizable.

In the same way ($p = 13$, $f(x) = x^4 - 52x^2 - 65x + 195$) the case $(B, 3)$ can be shown to exist. For here $f(x) \equiv x^4 \pmod{13}$ whence, since 13 is prime to the index of K, $13 = \mathfrak{p}^4$ in K.

The case $(B, 2)$ can also be liquidated if we take $p = 7$ and $f(x) = x^4 - 82x^2 + 175x + 15$. We find, in a manner similar to the previous cases that $7 = \mathfrak{p}^2$ in K. But to assure that $(B, 2)$ actually obtains, we must show that $\left(\frac{d}{p}\right) = \left(\frac{d}{7}\right) = +1$. The discriminant $D(f)$ of $f(x)$ differs from d only in square factors and, since $D(f)$ can be found, after a little calculation, to be $5 \cdot 7^2 \cdot 11 \cdot 127 \cdot 154{,}061$, $d = 5 \cdot 11 \cdot 127 \cdot 154{,}061$ and $\left(\frac{d}{7}\right) = +1$. Hence $(B, 2)$ exists.

---

[10] Cf. van der Waerden, *Moderne Algebra*, Bd. I, p. 191.

[11] Dedekind, *Werke*, Bd. I, p. 213.

At this point we wish to add the subgroup series inadvertently omitted from Table C. They are[12] as follows ($p = 2$):

$(C, 8)$: $G_Z = G_T = G_{V_1} = G_4$, $G_{V_2} = V_2$; $v_1 = 2$, $v_2 = 4$;

$(C, 9)$: $G_Z = C_4$, $G_T = G_{V_1} = V_2$; $v_1 = 2$;

$(C, 10)$: $G_Z = G_T = G_{V_1} = G_4$, $G_{V_2} = 1$; $v_1 = 1$;

$(C, 11)$: $G_Z = G_4$, $G_T = G_{V_1} = V_2$; $v_1 = 2$;

$(C, 12)$: $G_Z = G_4$, $G_T = G_{V_1} = C_2$; $v_1 = 1$.

The details involved will not be gone into; the mode of procedure is that used in Q.

The cases $(C, 2)$, $(C, 3)$, and $(C, 5)$ are all realizable since they correspond to different Hilbert subgroup series for B, and we can realize them by extending a properly chosen imaginary field B to a field N as in Theorem 5.

As in a few paragraphs above we can demonstrate the existence of $(C, 1)$ by a numerical example. We now take $f(x) = x^4 - 104x^2 - 84x + 70$, whose group, again, is S_4. Since $f(x) \equiv x^4 \pmod 2$ and since 2 is prime to the index of the field K generated by a root of $f(x) = 0$, $2 = \mathfrak{p}^4$ in K. Since only in $(C, 1)$ can such a decomposition occur (the cases $(C, 8)$ to $(C, 12)$ do not yield any such decomposition) the case $(C, 1)$ is hence realizable.

$(C, 12)$ is realizable also. We take $f(x) = x^4 + 105x^3 - 69x^2 - 14x + 70 \equiv x^2(x^2 + x + 1) \pmod 2$ and obtain $2 = \mathfrak{p}_1^2\mathfrak{p}_2$ in K, a quartic field with the symmetric group. Such a decomposition can occur, however, only in $(C, 12)$ as is easily seen.

In conclusion, the cases which have not been proved realizable or otherwise are the following: $(B, 4)$, $(B, 7)$–$(B, 9)$, $(C, 4)$, and $(C, 6)$–$(C, 11)$.

YALE UNIVERSITY.

---

[12] The cases $G_Z = G_T = G_{V_1} = C_4$, $G_{V_2} = V_2$, and $G_Z = G_T = G_{V_1} = G_4$, $G_{V_2} = C_2$ are impossible, as can be proved, using the theorems of Q.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ON CONGRUENCES INVOLVING BERNOULLI NUMBERS AND THE QUOTIENTS OF FERMAT AND WILSON

BY EMMA LEHMER

(Received December 1, 1937)

In this paper we give the residues of Bernoulli numbers modulo p^2 in terms of sums of like powers of numbers in arithmetical progression. It will be seen that the results obtained are as simple as those given for the residues of Bernoulli numbers modulo p by Glaisher[1] and later by an entirely different method by Vandiver.[2] We shall follow here the method of Glaisher which depends on Bernoulli polynomials of fractional arguments, rather than that of Vandiver, although both methods are capable of producing the results given below. We will apply these results to congruences involving Fermat's and Wilson's quotients, and generalize some results obtained by Friedmann and Tamarkin, Mirimanoff, Lerch, and Vandiver. We express Wilson's quotient modulo p in terms of $(p-2)^{\text{nd}}$ powers of numbers in arithmetical progression, and give some criteria for the divisibility of the Fermat's quotients q_2 and q_3 by p^2, and also some criteria for the 1st case of Fermat's Last Theorem, in terms of sums of reciprocals of numbers in arithmetical progression, and in terms of certain binomial coefficients.

We define the Bernoulli polynomial $B_\nu(x)$ by

$$B_\nu(x) = \sum_{r=0}^{\nu} \binom{\nu}{r} B_r x^{\nu-r}, \tag{1}$$

where, in turn, B_r are the Bernoulli numbers defined by

$$B_0 = 1, \quad B_1 = -\tfrac{1}{2}, \quad B_2 = \tfrac{1}{6}, \quad B_{2k+1} = 0 \text{ for } k > 0,$$

$$B_r = \sum_{i=0}^{r} \binom{r}{i} B_i.$$

If in the familiar difference equation

$$B_{\nu+1}(x+1) - B_{\nu+1}(x) = (\nu+1)x^\nu,$$

we let $x = (p - rn)/n$, $\left(r = 1, 2, \cdots, \left[\frac{p}{n}\right]\right)$,[3] where n and p are integers and $n < p$, and add all the resulting equations, we obtain after cancellation

$$\sum_{r=0}^{[p/n]} (p - rn)^\nu = \frac{n^\nu}{\nu+1}\left\{B_{\nu+1}\left(\frac{p}{n}\right) - B_{\nu+1}\left(\frac{s}{n}\right)\right\}, \tag{2}$$

[1] Quarterly Journal of Mathematics, v. 32, pp. 271–305, (1901).
[2] Proceedings Nat. Acad of Sci. v. 16, pp. 139–144, (1930).
[3] Here $[u]$ denotes as usual, the greatest integer not exceeding u.

where we have written s for the least positive residue of p modulo n. Setting $\nu = 2k$, and $x = p/n$, where p is an odd prime $> n$, in (1) we get the congruence

$$B_{2k}\left(\frac{p}{n}\right) \equiv B_{2k} + \frac{p^2}{n^2}\binom{2k}{2} B_{2k-2} \pmod{p^3}, \tag{3}$$

since $B_{2k-3} = 0$, and all the other terms are multiples of p^4 by a Bernoulli number, and therefore, by the von Staudt-Clausen theorem, are at least multiples of p^3. By the same theorem B_{2k-2} will contain p in the denominator only when $2k - 2$ is divisible by $(p - 1)$. It follows therefore that

$$B_{2k}\left(\frac{p}{n}\right) \equiv B_{2k} \pmod{p^2} \quad 2k \not\equiv 2 \pmod{p-1}. \tag{4}$$

Similarily we find for $\nu = 2k + 1$

$$B_{2k+1}\left(\frac{p}{n}\right) \equiv p\,\frac{2k+1}{n}\,B_{2k} \pmod{p^3} \quad 2k \not\equiv 2 \pmod{p-1}. \tag{5}$$

Substituting these results in (2), we obtain the following congruences

$$\sum_{r=1}^{[p/n]} (p - rn)^{2k-1} \equiv \frac{n^{2k-1}}{2k}\left\{B_{2k} - B_{2k}\left(\frac{s}{n}\right)\right\} \pmod{p^2}, \tag{6}$$

and

$$\sum_{r=1}^{[p/n]} (p - rn)^{2k} \equiv \frac{n^{2k}}{2k+1}\left\{\frac{2k+1}{n}\,pB_{2k} - B_{2k+1}\left(\frac{s}{n}\right)\right\} \pmod{p^3}, \tag{7}$$

where s is the least positive residue of $p \pmod{n}$, and $2k \not\equiv 2 \pmod{p-1}$.

Since

$$\sum_{r=1}^{[p/n]} (p - nr)^{\nu} \equiv (-1)^{\nu}\left\{n^{\nu}\sum_{r=1}^{[p/n]} r^{\nu} - p\nu n^{\nu-1}\sum_{r=1}^{[p/n]} r^{\nu-1}\right\} \pmod{p^2},$$

congruences (6) and (7) may be combined to give sums of like powers of numbers less than $[p/n]$. We can write

$$\sum_{r=1}^{[p/n]} r^{2k-1} \equiv \frac{1}{2k}\left\{B_{2k}\left(\frac{s}{n}\right) - B_{2k}\right\} - \frac{p}{n}\,B_{2k-1}\left(\frac{s}{n}\right) \pmod{p^2}, \tag{8}$$

and

$$\sum_{r=1}^{[p/n]} r^{2k} \equiv -\frac{B_{2k+1}\left(\frac{s}{n}\right)}{2k+1} + \frac{p}{n}\,B_{2k}\left(\frac{s}{n}\right) \pmod{p^2}. \tag{9}$$

Formulas (6) and (7) may be thought of as generalizations of Glaisher's results, while formulas (8) and (9) give generalizations of Vandiver's results whenever possible. Both sets of formulas depend on the evaluation of $B_r\left(\frac{s}{n}\right)$. This

evaluation has been effected in terms of Bernoulli numbers for ν even and $n = 1, 2, 3, 4$, and 6, while for ν odd, $B_\nu\left(\frac{s}{n}\right) = 0$ for $n = 1$ and 2, and is given in terms of Eulerian numbers for $n = 4$. In case $n = 3$ and 6, $B_\nu\left(\frac{s}{n}\right)$ is given in terms of what Glaisher calls I numbers,[4] but we shall not consider these cases here. Obviously both sets of formulas are equivalent modulo p. With respect to the modulus p^2 however the form (6) is superior to (8), since for $n > 2$, (8) involves a value of a Bernoulli polynomial not expressible in Bernoulli numbers. In case $n = 4$, both (7) and (9) give residues of Eulerian numbers. The values of $B_\nu\left(\frac{s}{n}\right)$ can be tabulated as follows.

$$B_\nu(1) = B_\nu, \quad \nu > 1.$$

$$B_\nu\left(\frac{1}{2}\right) = (1 - 2^{\nu-1})\frac{B_\nu}{2^{\nu-1}}, \quad \nu \text{ even}; \qquad B_\nu\left(\frac{1}{2}\right) = 0, \quad \nu \text{ odd}.$$

$$B_\nu\left(\frac{1}{3}\right) = B_\nu\left(\frac{2}{3}\right) = (1 - 3^{\nu-1})\frac{B_\nu}{2\cdot 3^{\nu-1}}, \quad \nu \text{ even}.$$

$$B_\nu\left(\frac{1}{4}\right) = B_\nu\left(\frac{3}{4}\right) = (1 - 2^{\nu-1})\frac{B_\nu}{2^{2\nu-1}}, \quad \nu \text{ even}.$$

$$B_\nu\left(\frac{1}{4}\right) = -B_\nu\left(\frac{3}{4}\right) = -\frac{\nu E_{\nu-1}}{4^\nu}, \quad \nu \text{ odd}.$$

$$B_\nu\left(\frac{1}{6}\right) = B_\nu\left(\frac{5}{6}\right) = (1 - 2^{\nu-1})(1 - 3^{\nu-1})\frac{B_\nu}{2^\nu\cdot 3^{\nu-1}}, \quad \nu \text{ even}.$$

These evaluations of $B_\nu\left(\frac{s}{n}\right)$ are well known.[5] Substituting these values in (6) we get at once for[6] $2k \not\equiv 2 \pmod{p-1}$

$$(10) \quad \sum_{r=1}^{p-1/2} (p - 2r)^{2k-1} \equiv (2^{2k} - 1)\frac{B_{2k}}{2k} \pmod{p^2},$$

$$(11) \quad \sum_{r=1}^{[p/3]} (p - 3r)^{2k-1} \equiv (3^{2k} - 1)\frac{B_{2k}}{4k} \pmod{p^2}, \quad p > 3,$$

$$(12) \quad \sum_{r=1}^{[p/4]} (p - 4r)^{2k-1} \equiv (2^{2k} - 1)(2^{2k-1} + 1)\frac{B_{2k}}{4k} \pmod{p^2}, \quad p > 3,$$

$$(13) \quad \sum_{r=1}^{[p/6]} (p - 6r)^{2k-1} \equiv (6^{2k-1} + 3^{2k-1} + 2^{2k-1} - 1)\frac{B_{2k}}{4k} \pmod{p^2}, \quad p > 5.$$

---

[4] Quarterly Jour. v. 28, p. 157.

[5] See for example N. E. Norlund, *Differenzenrechnung*, Berlin 1924, pp. 22 and 29. $B_\nu\left(\frac{s}{n}\right)$ has not been evaluated for any other values of n. The values 1, 2, 3, 4, and 6 can be characterized by the fact that their totient does not exceed two. It would be of interest to attempt the evaluation of $B_\nu\left(\frac{s}{n}\right)$ when the quotient of n is 4. Namely for $n = 5, 8, 10$ and 12.

[6] In the exceptional case $2k \equiv 2 \pmod{p-1}$ the congruences are true modulo p.

These results reduce modulo p to the congruences given by Glaisher for sums of negative powers of numbers in arithmetical progression, since for a prime to p

$$a^{-\nu} \equiv a^{p-\nu-1} \pmod{p}$$

by Fermat's theorem. Modulo p^2 however,

$$a^{-\nu} \equiv 2a^{p-\nu-1} - a^{2p-\nu-2} \pmod{p^2},$$

so that sums of negative odd powers of numbers in arithmetical progression can be obtained from the congruences (10) to (13) as a linear combination of two Bernoulli numbers whose subscripts differ by $(p-1)$. As these formulas do not simplify in general we shall not take space to write them down. We shall return later to the special case of sums of reciprocals in arithmetical progression.

First we will give the results of substituting $B_\nu\left(\frac{s}{n}\right)$ in (8) and (9) for $n = 1, 2$, and 4.

If $n = 1$, (8) and (9) are of course the same as (6) and (7). Moreover since $B_{2k}(1) = B_{2k}$ and $B_{2k+1}(1) = 0$, it follows that

$$\sum_{r=1}^{p-1} r^{2k+1} \equiv 0 \pmod{p^2}.$$

But we may go a step further and use (3) instead of (4) in (2) obtaining, $k > 0$

$$\sum_{r=1}^{p-1} r^{2k+1} \equiv p^2 \frac{2k+1}{2} B_{2k} \pmod{p^3}, \quad 2k \not\equiv 2 \pmod{p-1}, \tag{14}$$

while from (7)

$$\sum_{r=1}^{p-1} r^{2k} \equiv pB_{2k} \pmod{p^3}, \quad 2k \not\equiv 2 \pmod{p-1}. \tag{15}$$

This pair of congruences is a generalization of the familiar statement

$$\sum_{r=1}^{p-1} r^{\nu} \equiv pB_\nu \pmod{p^2}, \quad \nu \not\equiv 1 \pmod{p-1}. \tag{16}$$

In passing we remark that from these congruences we may also calculate sums of negative powers and get[7]

$$\sum_{r=1}^{p-1} 1/r^{2k} \equiv p(2k/2k+1)B_{p-1-2k} \pmod{p^2}$$

and

$$\sum_{r=1}^{p-1} 1/r^{2k-1} \equiv p^2 k[(1-2k)/(1+2k)]B_{p-1-2k} \pmod{p^3}, \quad 2k \not\equiv p-2 \pmod{p-1},$$

[7] See also Glaisher, Quar. Jour. v. 31, p. 231, (1900).

If $n = 2$, (8) becomes

$$(17) \qquad \sum_{r=1}^{(p-1)/2} r^{2k-1} \equiv (1 - 2^{2k}) \frac{B_{2k}}{2^{2k} k} \pmod{p^2}, \quad 2k \not\equiv 2 \pmod{p-1},$$

a result given by Mirimanoff.[8] Since $B_{2k+1}(\frac{1}{2}) = 0$, (9) can easily be shown to be true modulo p^3 so that

$$(18) \qquad \sum_{r=1}^{(p-1)/2} r^{2k} \equiv p(1 - 2^{2k-1}) \frac{B_{2k}}{2^{2k}} \pmod{p^3}, \quad 2k \not\equiv 2 \pmod{p-1},$$

while (7) gives

$$(19) \qquad \sum_{r=1}^{(p-1)/2} (p - 2r)^{2k} \equiv p 2^{2k-1} B_{2k} \pmod{p^3}, \quad 2k \not\equiv 2 \pmod{p-1}.$$

To obtain residues of Eulerian numbers modulo p^3, we may use (7) which gives for $2k \not\equiv 2 \pmod{p-1}$

$$(20) \qquad \sum_{r=1}^{[p/4]} (p - 4r)^{2k} \equiv (-1)^{(p-1)/2} \frac{E_{2k}}{4} + p 4^{2k-1} B_{2k} \pmod{p^3}.$$

This congruence can be combined with (19) to eliminate the Bernoulli number. Modulo p, (20) reduces to the expression given by Glaisher for E_{2k}. In the exceptional case $2k \equiv 2 \pmod{p-1}$ all the congruences given above modulo p^α, hold modulo $p^{\alpha-1}$ as can be seen from (4) and (5).

We shall now pause in our discussion to recall some of the fundamental properties of Fermat's quotient

$$(21) \qquad q_a = (a^{p-1} - 1)/p$$

and Wilson's quotient

$$(22) \qquad w_p = [(p-1)! + 1]/p.$$

It follows readily from the fundamental congruence

$$q_{ab} \equiv q_a + q_b \pmod{p}$$

that Fermat's and Wilson's quotients are connected by the relation

$$(23) \qquad \sum_{a=1}^{p-1} q_a \equiv w_p \pmod{p}.$$

If we now write (16) with $\nu = t(p-1)$ we obtain, since

$$r^{t(p-1)} \equiv 1 + ptq_r \pmod{p^2},$$

the relation

$$(24) \qquad p - 1 + ptw_p \equiv pB_{t(p-1)} \pmod{p^2},$$

---

[8] Jour. fur. Math., v. 115 pp. 295–300, (1895). In a recent paper, L'Enseignement Math. v. 36, pp. 228–235, (1937), Mirimanoff points out some errors is this paper.

a congruence which is usually given with $t = 1$. We will use it here with $t = 1$ and 2 to obtain by subtraction

$$w_p = B_{2(p-1)} - B_{p-1} \pmod{p} \tag{25}$$

a result which will be of use later. As a complementary result when $\nu \not\equiv 0 \pmod{p-1}$ we will need the following well know congruence[9]

$$\frac{B_{\nu+p-1}}{\nu+p-1} \equiv \frac{B_\nu}{\nu} \pmod{p}, \quad \nu \not\equiv 0 \pmod{p-1}. \tag{26}$$

We are now in a position to transform the sums (10) to (20) into sums involving Fermat's quotients by means of the relation

$$a^{\nu+p-1} - a^\nu = pa^\nu q_a .$$

To begin with, (14) becomes

$$\begin{aligned} \sum_{r=1}^{p-1} r^{2k+1} q_r &\equiv p(2kB_{2k+p-1} - (2k+1)B_{2k})/2 \\ &\equiv -pB_{2k} \pmod{p^2} \quad 2k \not\equiv 0, 2 \pmod{p-1} \end{aligned} \tag{27}$$

by (26); while (15) gives

$$\sum_{r=1}^{p-1} r^{2k} q_r = B_{2k+p-1} - B_{2k} \pmod{p^2}. \tag{28}$$

Congruences (27) and (28) generalize the following congruence

$$\sum_{r=1}^{p-1} r^\nu q_r \equiv \begin{cases} -B_\nu/\nu \pmod{p} & \nu \not\equiv 0 \pmod{p-1} \\ w_p & \nu \equiv 0 \pmod{p-1}, \end{cases}$$

obtained by Friedmann and Tamarkin[10] from the consideration of sums of greatest integers. (27) is also given by Nielsen.[11] Similarly we may obtain from (17), if $2k \not\equiv 0, 2 \pmod{p-1}$

$$\sum_{r=1}^{(p-1)/2} r^{2k-1} q_r \equiv \frac{1-2^{2k}}{2^{2k-1}} \left(\frac{B_{2k+p-1}}{2k+p-1} - \frac{B_{2k}}{2k} \right) \Big/ p - q_2 \frac{B_{2k}}{2^{2k}k} \pmod{p}. \tag{29}$$

This congruence was given by Mirimanoff[12] in the special case when B_{2k} is divisible by p. It might be of interest to notice that if $2^{2k} - 1$ is divisible by p, then

$$\sum_{r=1}^{(p-1)/2} r^{2k-1} q_r \equiv -q_2 \frac{B_{2k}}{k} \pmod{p}.$$

---

[9] See for example Bachmann, *Niedere Zahlentheorie* v. 2, p. 41.
[10] Crelle v. 137, p. 148, (1909).
[11] *Traité Elémentaire des Nombres de Bernoulli*, p. 368, Paris, 1923.
[12] Loc. cit.

Similar results may be written down for the remaining congruences, but we will confine ourselves from now on to the special cases of sums of $(p-2)^{nd}$ powers of numbers in arithmetical progression, which will give us residues of Wilson's quotient, and to sums of reciprocals which will lead to some criteria for the first case of Fermat's Last Theorem.

For $2k = p-1$, congruences (10) to (13) simplify by means of (24) written in the form

$$\frac{pB_{p-1}}{p-1} \equiv 1 + \frac{pw_p}{p-1} \equiv 1 - pw_p \pmod{p^2},$$

and we get

$$(30) \qquad \sum_{r=1}^{(p-1)/2} (p-2r)^{p-2} \equiv q_2(1-pw_p) \pmod{p^2},$$

$$(31) \qquad \sum_{r=1}^{[p/3]} (p-3r)^{p-2} \equiv q_3(1-pw_p)/2 \pmod{p^2},$$

$$(32) \qquad \sum_{r=1}^{[p/4]} (p-4r)^{p-2} \equiv \{3q_2(1-pw_p) + pq_2^2\}/4 \pmod{p^2},$$

$$(33) \qquad \sum_{r=1}^{[p/6]} (p-6r)^{p-2} \equiv \{(4q_2+3q_3)(1-pw_p) + pq_2q_3\}/12 \pmod{p^2},$$

while from (17)

$$(34) \qquad \sum_{r=1}^{(p-1)/2} r^{p-2} \equiv -2q_2(1-pw_p) + 2pq_2^2 \pmod{p^2}.$$

All these congruences (30) to (34) could be used for the calculation of the remainder of w_p modulo p. Congruence (33), having of course the least number of terms, is the most practical one to use. We would write (33) in the form

$$(35) \qquad pw_p(3q_3+4q_2) \equiv -12 \sum_{r=1}^{[(p-1)/6]} (p-6r)^{p-2} + (3q_3+4q_2+pq_2q_3) \pmod{p^2}.$$

For $p = 17$ for example $q_2 \equiv 98$, $q_3 \equiv 231$ (Mod 289), and there are only two terms in the sum, $5^{15} \equiv 41$ and $11^{15} \equiv 14 \pmod{289}$. Hence we have

$$14 \cdot 17\, w_{17} \equiv -12\,(41+14) + (115 + 103 + 17 \cdot 130) \pmod{289}$$

$$\equiv 2 \cdot 17 \pmod{289}$$

or

$$w_{17} \equiv 1/7 \equiv 5 \pmod{17}.$$

A formula equivalent to a combination of (12) and (13) modulo p was actually used by Vandiver for the calculations of the residues of Bernoulli numbers modulo p for $p < 600$ in his investigation of irregular primes in connection with Fermat's Last Theorem. In our case, however, it is debatable whether (34) is more practical for the calculation of the remainders of w_p modulo p for large values of p, than the factorial definition of w_p. w_p has been computed modulo p by Beeger[13] for $p < 300$ from the factorial definition, and for $p < 211$ by the author,[14] using (24), and the recently extended table of Bernoulli numbers.[15] The errors in Beeger's table were given later.[16] These tables show that for $p < 300$, $w_p \equiv 0 \pmod{p}$ only for $p = 5$ and 13. The above formulas give criteria for the divisibility of w_p by p in terms of q_2 and q_3 and sums of $(p-2)^{nd}$ powers, but do not throw any light on the problem of finding a $p > 13$ for which $w_p \equiv 0 \pmod{p}$. A set of congruences similar to (30)–(34) could be written down for $2k = 2(p-1)$, or in fact more generally for $2k = t(p-1)$, and since $a^{2(p-1)} - a^{p-1} = pq_a$ we obtain a set of congruences for w_p in terms of q's whose subscripts are in arithmetical progression.

$$\sum_{r=1}^{(p-1)/2} \frac{1}{r} q_{p-2r} \equiv 2q_2 w_p - q_2^2 \pmod{p}. \tag{36}$$

$$\sum_{r=1}^{[p/3]} \frac{1}{r} q_{p-3r} \equiv \frac{3}{2} q_3 w_p - \frac{3}{4} q_3^2 \pmod{p}, \quad p > 3. \tag{37}$$

$$\sum_{r=1}^{[p/4]} \frac{1}{r} q_{p-4r} \equiv 3q_2 w_p - \frac{5}{2} q_2^2 \pmod{p}, \quad p > 3. \tag{38}$$

$$\sum_{r=1}^{[p/6]} \frac{1}{r} q_{p-6r} \equiv \left(2q_2 + \frac{3}{2} q_3\right) w_p - \left(q_2^2 + \frac{3}{4} q_3^2 + \frac{1}{2} q_2 q_3\right) \pmod{p}, \quad p > 5. \tag{39}$$

and

$$\sum_{r=1}^{(p-1)/2} \frac{1}{r} q_r \equiv 2q_2 w_p + q_2^2 \pmod{p}. \tag{40}$$

Formula (39) might seem at first sight preferable to (34) for the calculation of w_p since it involves the same number of terms and gives w_p modulo p directly rather than pw_p modulo p^2. The calculation of q_a however involves the calculation of a^{p-1} modulo p^2 so there is really no saving of time or labor, but instead a loss of a valuable check modulo p which (34) affords.

We now return to the sums of reciprocals of numbers in arithmetical progression. These can be now obtained by subtracting the corresponding for-

---

[13] Mess. of Math. v. 49, pp. 177–8, (1920).
[14] Amer. Math. Monthly, v. 44, pp. 237–8, (1937).
[15] Duke Jour. v. 2, pp. 462–4, (1936).
[16] Amer. Math. Monthly, v. 44, p. 462, (1937).

mulas in the sets (30) to (34) and (36) to (40) since $\frac{1}{a} = a^{p-2} - p\frac{q_a}{a}$. In this way we obtain

$$(41)\qquad \sum_{r=1}^{(p-1)/2} \frac{1}{p-2r} \equiv q_2 - pq_2^2/2 \pmod{p^2}$$

$$(42)\qquad \sum_{r=1}^{[p/3]} \frac{1}{p-3r} \equiv \frac{q_3}{2} - \frac{pq_3^2}{4} \pmod{p^2}, \quad p > 3.$$

$$(43)\qquad \sum_{r=1}^{[p/4]} \frac{1}{p-4r} \equiv \frac{3}{4}q_2 - \frac{3}{8}pq_2^2 \pmod{p^2}, \quad p > 3.$$

$$(44)\qquad \sum_{r=1}^{[p/6]} \frac{1}{p-6r} \equiv \frac{1}{4}q_3 + \frac{1}{3}q_2 - p\left(\frac{1}{8}q_3^2 + \frac{1}{6}q_2^2\right). \pmod{p^2}, \quad p > 5,$$

and

$$(45)\qquad \sum_{r=1}^{(p-1)/2} \frac{1}{r} \equiv -2q_2 + pq_2^2 \pmod{p^2}.$$

Congruences (41) and (42) were given modulo p by Lerch,[17] while (41) and (43) were obtained modulo p by Glaisher.[18] Vandiver[19] stated (41) as follows:

$$q_2 \equiv 0 \pmod{p^2} \textit{ if and only if } \sum_{a=1}^{(p-1)/2} \frac{1}{p-2a} \equiv 0 \pmod{p^2}.$$

We can combine (41) with (43) to give a slightly stronger condition:

$$q_2 \equiv 0 \pmod{p^2} \textit{ if and only if } \sum_{a=1}^{[p/4]} \frac{1}{p-4a} \equiv \sum_{a=1}^{[p/4]} \frac{1}{p-2-4a} \equiv 0 \pmod{p^2},$$

while a similar condition for $q_3 \equiv 0 \pmod{p^2}$ can be read out of (42).

It follows from (41)–(43) that the assumption $q_2 \equiv q_3 \equiv 0 \pmod{p}$ implies

$$\sum_{r=1}^{[p/n]} \frac{1}{r} \equiv 0 \pmod{p}, \quad \text{for } n = 2, 3, 4, \text{ and } 6.$$

But $q_2 \equiv 0 \pmod{p}$ and $q_3 \equiv 0 \pmod{p}$ are respectively the criteria of Wieferich and Mirimanoff for the first case of Fermat's Last Theorem. We can therefore transform a combination of these criteria as follows:

The equation

$$(46)\qquad x^p + y^p + z^p = 0$$

has no solutions in integers x, y, z prime to p unless

$$\sum_{r=1}^{[p/n]} \frac{1}{r} \equiv 0 \pmod{p}, \quad n = 2, 3, 4, \text{ and } 6.$$

---

[17] Math. Annalen, v. 60, pp. 471–490, (1905).
[18] Quarterly Jour. v. 32, pp. 1–27, (1901).
[19] Annals of Math. v. 18, pp. 112, (1917).

Vandiver[20] derived the same condition in the case $n = 5$ using Kummer's criteria. Combining the conditions for $n = 5$ and 6 for instance we can state that $\sum_{r=[p/6]+1}^{[p/5]} \frac{1}{r} \equiv 0 \pmod{p}$ is a criterion for the first case of Fermat's Last Theorem.[21]

Vandiver[22] has also proved that if (46) has a solution in integers x, y, z, prime to p, then

$$\sum_{r=1}^{[p/3]} \frac{1}{r^2} \equiv 0 \pmod{p}. \tag{47}$$

It has been shown by Schwindt,[23] and can also be made to follow from (7) with $2k = p - 3$, and $n = 3$ and 6, that

$$5 \sum_{r=1}^{[p/3]} \frac{1}{r^2} \equiv \sum_{r=1}^{[p/6]} \frac{1}{r^2} \pmod{p}$$

so that

$$\sum_{r=1}^{[p/6]} \frac{1}{r^2} \equiv 0 \pmod{p} \tag{48}$$

is also a criterion for the first case of Fermat's Last theorem. These last two criteria can be restated in terms of Glaisher's I-numbers referred to above. If it were proved that $\sum_{r=1}^{[p/4]} \frac{1}{r^2} \equiv 0 \pmod{p}$ is also a criterion for the first case of Fermat's Last Theorem, then a criterion could be given in terms of Eulerian numbers, since from (20)

$$\sum_{r=1}^{[p/4]} \frac{1}{r^2} \equiv (-1)^{(p-1)/2} 4E_{p-3} \pmod{p}.$$

It follows further since $\frac{1}{p - nr} \equiv -\frac{1}{nr} - p\frac{1}{n^2r^2} \pmod{p^2}$, that, by substituting the criteria (47) and (48) together with $q_2 \equiv q_3 \equiv 0 \pmod{p}$, into (42) and (44), we obtain

$$\sum_{r=1}^{[p/3]} \frac{1}{r} \equiv -3 \sum_{r=1}^{[p/3]} \frac{1}{p - 3r} \equiv -\frac{3}{2} q_3 \pmod{p^2}$$

and

$$\sum_{r=1}^{[p/6]} \frac{1}{r} \equiv -6 \sum_{r=1}^{[p/6]} \frac{1}{p - 6r} \equiv -\frac{3}{2} q_3 - 2q_2 \pmod{p^2},$$

---

[20] Jour. für Math. v. 144, pp. 314–318, (1914).

[21] Frobenius, Berlin Sitz. 1914, pp. 653–81, gave similar criteria in terms of

$$s_k = \Sigma\, 1/r, \qquad ((k-1)p/n < r < kp/n)$$

for $n = 7$ and $n = 12$ (p. 676). Other criteria in terms of linear combinations of s_k can be derived from his results for $n \leqq 26$.

[22] Annals of Math. v. 26, pp. 88–94, (1924).

[23] Jahrber. Deutsche Math. Ver. v. 43, pp. 229–231, (1933–4).

if Fermat's equation has a solution in integers x, y, z, prime to p. Moreover since $\sum_{r=1}^{p-1} \frac{1}{r} \equiv 0 \pmod{p^2}$,

$$\sum_{r \equiv -p \pmod 3} \frac{1}{r} \equiv 0 \pmod{p^2}$$

is also a criterion for the first case of Fermat's Last Theorem.

In conclusion we apply our congruences (41)–(45) to the problem of finding the residues modulo p^2 and p^3 of certain binomial coefficients. Since[24]

$$\binom{p-1}{k} \equiv (-1)^k \left\{1 - p\sum_{r=1}^{k} \frac{1}{r} + \frac{p^2}{2}\left(\sum_{r=1}^{k} \frac{1}{r}\right)^2 - \frac{p^2}{2}\sum_{r=1}^{k} \frac{1}{r^2}\right\} \pmod{p^3}$$

it follows from (45) that[25] for $p > 3$

$$\binom{p-1}{(p-1)/2} \equiv (-1)^{(p-1)/2}\{1 - p(-2q_2 + pq_2^2) + 2p^2q_2^2\} \pmod{p^3}$$

$$\equiv (-1)^{(p-1)/2}(1 + pq_2)^2 \equiv (-1)^{(p-1)/2}\, 4^{p-1} \tag{49}$$

while using (41)–(44) modulo p we obtain for $p > 3$

$$\binom{p-1}{[p/3]} \equiv (-1)^{[p/3]}\left(\frac{3}{2}pq_3 + 1\right) \equiv (-1)^{[p/3]}(3^p - 1)/2 \pmod{p^2} \tag{50}$$

$$\binom{p-1}{[p/4]} \equiv (-1)^{[p/4]}(3pq_2 + 1) \equiv (-1)^{[p/4]}(3\cdot 2^{p-1} - 2) \pmod{p^2} \tag{51}$$

and for $p > 5$

$$\binom{p-1}{[p/6]} \equiv (-1)^{[p/6]}\left(2pq_2 + \frac{3}{2}pq_3 + 1\right) \equiv (-1)^{[p/6]}(2^{p+1} + 3^p - 5)/2 \pmod{p^2}. \tag{52}$$

It follows from these congruences that some of the criteria for Fermat's Last Theorem given above can be restated in terms of binomial coefficients as follows

The Fermat Equation (46) *has no solutions in integers x, y, z, prime to p unless*

$$\binom{p-1}{[p/n]} \equiv (-1)^{[p/n]} \pmod{p^2}$$

for $n = 2, 3, 4, 5,$ and 6.

Furthermore it follows from the criteria (47) and (48) that if Fermat's equation has a solution in integers x, y, z prime to p, then (50) and (52) are true modulo p^3.

BETHLEHEM, PA.

[24] This follows from the identity

$$\binom{p-1}{k} = (-1)^k\{(1 - pS_1 + p^2S_2 - + \cdots + (-1)^k p^k S_k)\}$$

where S_ν is the ν-th elementary symmetric function of $\left(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \cdots, \frac{1}{k}\right)$.

[25] This is equivalent to a result given by Nielsen: K. Danske Vidensk. Selsk. Skrifter, (7), v. 10, (1913), p. 353, formula (9).

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

SOME STUDIES ON REGULAR REPRESENTATIONS, INDUCED REPRESENTATIONS AND MODULAR REPRESENTATIONS

BY TADASI NAKAYAMA

(Received December 1, 1937)

Introduction

In their joint note B.N.R.[1] R. Brauer and C. Nesbitt studied the structure of the regular representations of algebras, and using the results they developed in B. N. M.[2] a very interesting theory of modular representations of finite groups. In the present note I adhere throughout to these two papers. In (1) some of their results concerning regular representations will be studied from the view point of the structural theory of algebras.[3] In (2) we extend those results to get some relations between an algebra and its subalgebra, which we can consider as an extension of Frobenius' theorem on induced characters[4] to the non-semi-simple case. In (3) we shall study induced modular characters of finite groups, and obtain from the result an application to representations of finite groups by semilinear transformations.[5] Furthermore, in (4) we shall see that the fundamental relation between Cartan invariants and decomposition numbers, which plays a principal rôle in the theory of modular representations, can be derived also from our generalized Frobenius theorem.

1. We consider an (associative) ring A with a unit element 1 and satisfying the minimum and maximum conditions for left and right ideals. Let N be the radical of A, whence the residue class ring A/N is semisimple. Let

$$\bar{A} = A/N = \bar{A}\bar{e}_1 + \bar{A}\bar{e}_2 + \cdots + \bar{A}\bar{e}_l, \qquad \bar{e}_i \text{ idempotent } (\epsilon \bar{A})$$

be a decomposition of $\bar{A} = A/N$ into direct sum of simple left ideals $\bar{A}\bar{e}_i$. Then $\bar{A} = \bar{e}_1\bar{A} + \bar{e}_2\bar{A} + \cdots + \bar{e}_l\bar{A}$ is a decomposition of $\bar{A}$ into direct sum of

---

[1] R. Brauer and C. Nesbitt, *On regular representations*, Proc. Nat. Acad. Sci. **23** (1937) ——— referred to as B.N.R.

[2] R. Brauer and C. Nesbitt, *On the modular representations of groups of finite order*. I., Univ. Toronto Studies, Math. Series **4** (1937) ——— referred to as B.N.M. Cf. also R. Brauer, *Darstellungen von Gruppen in Galoisschen Feldern*, Act. sci. et industr. **195** (1935).

[3] See for example M. Deuring, *Algebren*, Ergebn. der Math. **4** (1935).

[4] G. Frobenius, *Über Relationen zwischen den Characteren einer Gruppe und denen ihrer Untergruppen*, Sitz.-ber. Berlin 1898. Cf. also J. Levitzki, *Über vollständig reduzible Ringe und Unterringe*, Math. Zeitschr. **33** (1931).

[5] T. Nakayama and K. Shoda, Jap. Journ. Math. **12** (1936). Cf. also the last section of A. H. Clifford, *Representations induced in an invariant subgroup*, Ann. Math. **38** (1937).

simple right ideals. $\bar{A}\bar{e}_i$ and $\bar{A}\bar{e}_j$ are (operator-) isomorphic if and only if $\bar{e}_i\bar{A}$ and $\bar{e}_j\bar{A}$ are so.

There exist l mutually orthogonal idempotent elements e_i in A such that $e_i \in \bar{e}_i$ and $e_1 + e_2 + \cdots + e_l = 1$. Then

$$A = Ae_1 + Ae_2 + \cdots + Ae_l = e_1A + e_2A + \cdots + e_lA,$$

and here Ae_i and e_iA are (directly) indecomposable left and right ideals of A respectively. Ae_i and Ae_j (e_iA and e_jA) are isomorphic (with respect to A) if and only if $\bar{A}\bar{e}_i$ and $\bar{A}\bar{e}_j$ ($\bar{e}_i\bar{A}$ and $\bar{e}_j\bar{A}$) are so (with respect to $\bar{A}$, and then also with respect to A). Hence Ae_i and Ae_j are isomorphic if and only if e_iA and e_jA are so. All these are well known.[6] In particular, the statement last but one can be seen as follows; Assume $\bar{A}\bar{e}_i$ and $\bar{A}\bar{e}_j$ are isomorphic. Then $(e_i + e_j)A(e_i + e_j)$ is a primary (= primaer) ring and there exist such elements e_{ij}, e_{ji} that $e_ue_{uv} = e_{uv}e_v = e_{uv}$, $e_{uv}e_{vu} = e_u$ for $u = i$, $v = j$ or $u = j$, $v = i$, hence $Ae_j = Ae_ie_{ij}$ is obviously isomorphic to Ae_i.

THEOREM 1. *$Ne_i = Ae_i \frown N$ is a maximal subideal of the left ideal Ae_i of A, and Ae_i has no other maximal subideal. Ae_i/Ne_i is isomorphic to $\bar{A}\bar{e}_i$ (with respect to A). Moreover, there is no other subideal M of Ae_i (than this Ne_i) such that Ae_i/M is completely reducible (as an A-left-module).*

PROOF. $Ae_i/Ne_i = Ae_i/Ae_i \frown N \cong (Ae_i, N)/N = \bar{A}\bar{e}_i$. Hence Ne_i is a maximal subideal, for $\bar{A}\bar{e}_i$ is simple. Let M be any maximal subideal of Ae_i. Then Ae_i/M is simple, whence $N \cdot (Ae_i/M) = 0$, that is, $NAe_i = Ne_i \subsetneq M$. This implies $M = Ne_i$.

To prove the second part, let Ae_i/M be completely reducible and let $Ae_i/M = L_1/M + \cdots + L_s/M$ be its decomposition into a direct sum of simple submoduli. Then $Ae_i/(L_2, \cdots, L_s) \cong L_1/M$ is simple, hence $(L_2, \cdots, L_s) = Ne_i$ by the first part of the theorem. If $s \geqq 2$, then $(L_1, L_3, \cdots, L_s) = Ne_i$ similarly, whence $Ae_i = (L_1, L_2, \cdots, L_s) = Ne_i$, which is obviously a contradiction. Therefore $s = 1$ and $M = Ne_i$.

THEOREM 2. *The composition length (the length of a composition series[7]) of the e_jAe_j[8]-left-module e_jAe_i[9] is equal to the number of the factor groups (residue class moduli) isomorphic to $\bar{A}\bar{e}_j$ which appear in a composition factor group series[10] of the left ideal Ae_i of A. (This number we denote by c_{ji}.)*

PROOF: LEMMA. *Let T be a submodule of the e_jAe_j-left-module e_jAe_i, and let T' be a proper submodule of T. Then $AT \supset AT'$. ($\supset$ means $\supseteq$ and $\neq$.) Moreover, if M is a maximal subideal of the left ideal AT of A, then AT/M is isomorphic to $\bar{A}\bar{e}_j$.*

---

[6] See Deuring, l. c. Chap. II, §6.

[7] Its existence can be seen from the proof below.

[8] e_jAe_j is a so-called completely primary ring.

[9] e_jAe_i is isomorphic to the module of homomorphisms of Ae_j into Ae_i.

[10] If $M = M_0, M_1, \cdots, M_t = 0$ is a composition series of a group M, then we shall call $M/M_1, M_1/M_2, \cdots, M_{t-1}/0$ a composition factor group series of M. An individual $M_\nu/M_{\nu+1}$ is called a composition factor group of M.

PROOF. $AT \supset AT'$ is obvious, since $AT = AT'$ implies $T = e_iAT = e_iAT' = T'$. Let M be a maximal subideal of AT. AT/M is simple, and $Ae_i \cdot (AT/M) = (Ae_iAe_iT, M)/M = AT/M$. This implies $AT/M \cong \bar{A}\bar{e}_j$, as is well known.

LEMMA. *Let* $Ae_i \supseteq M_1 \supset M_2$ *and let* M_1/M_2 *be simple and* $\cong \bar{A}\bar{e}_j$. *Then* $(e_jAe_i \supseteq)\ e_jM_1 \supset e_jM_2$.

PROOF. $M_1/M_2 \cong \bar{A}\bar{e}_j$ implies $(Ae_jM_1, M_2) = M_1$. Hence $Ae_jM_1 \not\subseteq M_2$. Therefore $e_jM_1 \supset e_jM_2$, since $e_jM_1 = e_jM_2$ implies $Ae_jM_1 = Ae_jM_2 \subseteq M_2$.

From these two lemmas, which prove two inequalities of different directions concerning the two numbers stated in our theorem, we have certainly our theorem.

REMARK. As one can easily see, we can replace in this theorem e_iAe_i and Ae_i by $e_iA'e_i$ and $A'e_i$ respectively, where A' is any two-sided ideal of A.

COROLLARY. *The sum of the composition lengths of the* $\bar{e}_j\bar{A}\bar{e}_j$ $(= e_jAe_j/e_jNe_j)$[11]*-left-moduli* e_jAe_i/e_jNe_i, $e_jNe_i/e_jN^2e_i$, $\cdots$ *is equal to the number of the factor groups isomorphic to* $\bar{A}\bar{e}_j$ *which appear in a composition factor group series of the left ideal* Ae_i.

This is immediate.

COROLLARY. *The number of the factor groups isomorphic to* $\bar{A}\bar{e}_j$ *which appear in a composition factor group series of A-left-module A is equal to the composition length of the* e_jAe_j*-left-module* e_jA.

PROOF. Both the numbers are equal to $\sum_i c_{ji}$.

We have of course the (right, left) duals of the above theorems. Let $\bar{c}_{ij}$ be the number of the factor groups $\cong \bar{e}_i\bar{A}$ in a composition factor group series of the right ideal e_jA. $\bar{c}_{ij}$ is at the same time the composition length of the e_iAe_i-right-module e_jAe_i.

THEOREM 3. *If A is an algebra (of a finite rank) over a field K, then*

$$c_{ji} \cdot (\text{Rank } \bar{e}_j\bar{A}\bar{e}_j) = \bar{c}_{ij} \cdot (\text{Rank } \bar{e}_i\bar{A}\bar{e}_i).$$

If moreover K is algebraically closed, then certainly $c_{ji} = \bar{c}_{ij}$.

PROOF. This follows immediately from calculation of the rank of e_jAe_i (over K). In case of an algebraically closed K both the division algebras $\bar{e}_i\bar{A}\bar{e}_i$ and $\bar{e}_j\bar{A}\bar{e}_j$ are isomorphic to K.

Our Theorem 1 and its dual correspond to the first two statements in B. N. R. concerning the first (last) irreducible and largest completely reducible constituents of an indecomposable part of the second (first) regular representation.[12]

It is easy to see that if Ae_i and $Ae_{i'}$ are isomorphic and if Ae_j and $Ae_{j'}$ are so too, then $c_{ji} = c_{j'i'}$. Hence the following $c^{(\lambda)(\kappa)}$ are more convenient: Let k be the number of Ae_i mutually non-isomorphic, and take $e^{(1)}, e^{(2)}, \cdots, e^{(k)}$ from $e_1, e_2, \cdots, e_l$ so that $Ae^{(1)}, Ae^{(2)}, \cdots, Ae^{(k)}$ are all mutually non-isomorphic.

---

[11] $\bar{e}_j\bar{A}\bar{e}_j$ is a skew field (= Schiefkoerper).

[12] For the connection between the representation theory and the structure of algebras see the original papers of E. Noether; Cf. also Deuring, l. c. and B. L. van der Waerden, *Gruppen von linearen Transformationen*, Ergebn. der Math. **4** (1935).

We put $c^{(\lambda)(\kappa)} = c_{ji}$ with $Ae_i \cong Ae^{(\kappa)}$ and $Ae_j \cong Ae^{(\lambda)}$. We define $\bar{c}^{(\kappa)(\lambda)}$ in a similar manner. Then these $c^{(\lambda)(\kappa)}$ and $\bar{c}^{(\kappa)(\lambda)}$, which coincide in case of an algebra over an algebraically closed field, correspond to *Cartan invariants* $c_{\lambda\kappa}$ of B. N. R. The formulas in B. N. R. concerning the Cartan invariants and the degrees of irreducible and indecomposable components of the regular representations can be derived from the above considerations. The third characterization of the $c_{\lambda\kappa}$ there is an immediate consequence of the characterization of c_{ji} by e_jAe_i. As for the characterization of Cartan invariants by means of the determinant $(S(x) + R(y)')$, it relates closely to the structure of the composition series of two-sided ideals of A. In our general case we have

THEOREM 4. *Let $z_{\lambda\kappa}$ denote the number of those factor groups in a composition factor group series of the A-two-sided module A, which are, considered as A-left-moduli, a direct sum of simple left-moduli isomorphic to $\bar{A}\bar{e}^{(\lambda)}$ and which are at the same time, considered as A-right-moduli, a direct sum of simple right-moduli isomorphic to $\bar{e}^{(\kappa)}\bar{A}$. Then $z_{\lambda\kappa}$ is equal to the composition length of the $e^{(\lambda)}Ae^{(\lambda)} - e^{(\kappa)}Ae^{(\kappa)}$-double-module $e^{(\lambda)}Ae^{(\kappa)}$. Hence $z_{\lambda\kappa} \leqq c^{(\lambda)(\kappa)}$ and $\leqq c^{(\kappa)(\lambda)}$.* (The composition length of the A-two-sided module A, that is, the length of a composition series of two-sided ideals of A, is $\sum z_{\lambda\kappa}$. In case of an algebra over an algebraically closed field the inequalities turn to equalities.)

PROOF. This we can prove in a similar manner as Theorem 2, the process being only doubled.

Let us next consider the decomposition of A into a direct sum of (directly) indecomposable two-sided ideals. We say, after B.N.M. §6, e_i and e_j belong to the same *block* if there exists a sequence $e_i, e_r, \cdots, e_s, e_j$ such that any two neighboring e_u and e_v define left ideals Ae_u and Ae_v which have at least one composition factor group in common (up to isomorphism).

THEOREM 5. *Let $\mathfrak{B}_1, \mathfrak{B}_2, \cdots, \mathfrak{B}_t$ be such blocks, and let A_ρ be the (direct) sum of the Ae_i with $e_i \in \mathfrak{B}_\rho$. Then $A = A_1 + A_2 + \cdots + A_t$ is the unique decomposition of A into a direct sum of indecomposable two-sided ideals.*

PROOF. First we prove that these A_ρ are two-sided; it suffices to prove that the A_ρ are right ideals. For this purpose it is sufficient, as one may easily see, to prove that $Ae_iA \frown Ae_j = 0$ if e_i and e_j belong to different blocks. Let us assume $Ae_iA \frown Ae_j \neq 0$, that is, $Ae_iAe_j \neq 0$. Then $e_iAe_j \neq 0$, whence $c_{ij} \neq 0$. Then e_i and e_j certainly belong to the same block.

Let next e_i and e_j belong to the same block, and let $e_i, e_r, \cdots, e_s, e_j$ be a sequence with the property stated in the above definition. Let e_u, e_v be any two neighboring elements in the sequence, and let both the Ae_u and Ae_v have a composition factor group isomorphic to $\bar{A}\bar{e}_h$. Then $e_hAe_u \neq 0$ and $e_hAe_v \neq 0$. Hence e_u and e_v must be contained in the same summand in any direct decomposition of A into two-sided ideals, since $e_hAe_u \subseteq Ae_hA$, Ae_uA and $e_hAe_v \subseteq Ae_hA$, Ae_vA. This holds for any neighboring e_u, e_v, and therefore all members of the sequence must belong to the same summand in any decomposition of A into two-sided ideals. This completes the proof.

2. Let $A, l, k, e_i, e^{(\kappa)}$ have the same meaning as in §1. We consider besides A another ring B containing A and having the unit element 1 in common with A. We assume furthermore the minimum and maximum conditions for both the A-left and A-right submoduli of B. Let $E_1, E_2, \cdots, E_n$ be mutually orthogonal idempotent elements in B such that $\sum E_q = 1$ and BE_q are indecomposable. Let $\bar{B}$ be the residue class ring of B with respect to its radical, and let $\bar{E}_q$ be the residue class containing E_q. In a similar manner as in §1, we have

THEOREM 6. *The composition length of the e_iAe_i-left-module e_iBE_q is equal to the number of the factor groups isomorphic to $\bar{A}\bar{e}_i$ which appear in a composition factor group series of the A-left-module BE_q. (This number we denote by b_{iq}.)*

PROOF. We have only to prove the following relations which can be verified in quite a similar way as in (1); 1) If $T \supset T'$ are two submoduli of the e_iAe_i-left-module e_iBE_q, then $AT \supset AT'$ and $AT/M \cong \bar{A}\bar{e}_i$ for any maximal submodule M of the A-left-module AT. 2) If $M_1 \supset M_2$ are two submoduli of BE_q with $M_1/M_2 \cong \bar{A}\bar{e}_i$, then $e_iM_1 \supset e_iM_2$.

We have moreover

THEOREM 7. *The composition length of the E_qBE_q-right-module e_iBE_q is equal to the number of the factor groups isomorphic to $\bar{E}_q\bar{B}$ which appear in a composition factor group series of the right ideal e_iB of B (induced by the right ideal e_iA of A). (This number we denote by $\tilde{a}_{qi}$.)*

THEOREM 8. *If A and B are algebras over a field K, then*

$$b_{iq} \cdot (\text{Rank } \bar{e}_i\bar{A}\bar{e}_i) = \tilde{a}_{qi} \cdot (\text{Rank } \bar{E}_q\bar{B}\bar{E}_q).$$

If moreover K is algebraically closed, then $b_{iq} = \tilde{a}_{qi}$.

We have of course the (right, left) duals to these theorems, and correspondingly we have the numbers $\tilde{b}_{iq}$ and a_{qi}. Theorem 8 and its dual show $\tilde{b}_{iq}/b_{iq} = a_{qi}/\tilde{a}_{qi}$ in case of algebras.

Now, let us take $E^{(1)}, E^{(2)}, \cdots, E^{(m)}$ from $E_1, E_2, \cdots, E_n$ such that $BE^{(1)}$, $BE^{(2)}, \cdots, BE^{(m)}$ are all mutually non-isomorphic (with respect to B) and every BE_q is isomorphic to one of them. Define $b^{(\kappa)(\alpha)}$, $\tilde{b}^{(\kappa)(\alpha)}$, $a^{(\alpha)(\kappa)}$, $\tilde{a}^{(\alpha)(\kappa)}$ in the same way as we defined $c^{(\lambda)(\kappa)}$, $\tilde{c}^{(\kappa)(\lambda)}$.

Consideration corresponding to Theorem 4 is also possible, and that leads us in case of algebras over an algebraically closed field to an extension of the formula concerning $\det(S(x) + R(y)')$.

The above Theorem 8 (together with the definitions of b_{iq} and $\tilde{a}_{qi}$) can be considered as an extension of Frobenius' theorem on induced characters to the *non-semisimple case*. Applying it to modular representations of finite groups, we have the following (3).

3. Let G' be a finite group and let G be its sugroup. We consider representations of G' and G in an algebraically closed field K. Our concern lies of course in the *modular* case. Let p be the characteristic of K. Let $F_1, F_2, \cdots, F_k$ and $F'_1, F'_2, \cdots, F'_m$ be distinct irreducible representations (in K) of G and G'.

Let $U_1, U_2, \cdots, U_k$ and $U_1', U_2', \cdots, U_m'$ be corresponding indecomposable parts of the (first, for example), regular representations of G, G' respectively. (Here we understand the correspondence between F_κ and $U_\kappa(F_\alpha'$ and $U_\alpha')$ in the sense of Theorem 1. In virtue of the fact that group rings are symmetric (in the sense of B.N.R.) $U_\kappa(U_\alpha')$ has $F_\kappa(F_\alpha')$ as its first and its last constituent). The characters of F_κ, U_κ, F_α', U_α' we denote by $\chi^{(\kappa)}$, $\varphi^{(\kappa)}$, $\mathrm{X}^{(\alpha)}$ and $\Phi^{(\alpha)}$ respectively. We understand these characters in the sense of B.N.M. (5); they are complex numbers and are defined only for the elements with an order prime to p.

Furthermore, we define induced representations of G' as usual in the following way: Let $x \to T(x)$ be a representation of G. Let $G' = s_1G + s_2G + \cdots + s_tG$ be the left coset decomposition of G' with respect to G. Then $y \to (T(s_u y s_v^{-1}))_{uv}$, where $T(z)$ is defined to be the zero matrix for z not contained in G, forms a representation of G', and is called the representation of G' induced from the representation T of G.

Theorem 8, applied to the group rings of G and G' over K, gives then

THEOREM 9. *Let*

$$\Phi^{(\alpha)}(x) = \sum_{\kappa=1}^{k} b^{(\kappa)(\alpha)} \chi^{(\kappa)}(x) \qquad \text{for } x \in G,$$

then

$$\Psi^{(\kappa)}(y) = \sum_{\alpha=1}^{m} b^{(\kappa)(\alpha)} \mathrm{X}^{(\alpha)}(y) \qquad \text{for } y \in G'$$

with the same $b^{(\kappa)(\alpha)}$, where $\Psi^{(\kappa)}$ denotes the character of G' induced from $\varphi^{(\kappa)}$, that is, the character of the representation of G' induced from U_κ. (x and y are elements of G and G' with order prime to p.)

Now, by B.N.M. k and m are equal to the number of those classes of conjugate elements in G and G' respectively which contain elements of an order prime to p. Let $\mathfrak{K}_1, \mathfrak{K}_2, \cdots, \mathfrak{K}_k$ and $\mathfrak{K}_1', \mathfrak{K}_2', \cdots, \mathfrak{K}_m'$ denote such classes of conjugate elements in G and G' respectively. Since the matrix $(\Phi^{(\alpha)}(\mathfrak{K}_\beta'))_{\alpha\beta}$ has the rank m, the maximal number of linearly independent characters from the m characters $\Phi^{(\alpha)}(x)$ $(x \in G)$ of G is equal to the number of those $\mathfrak{K}_\alpha'$ which contain an element of G. Let h denote this number. From the first equality in Theorem 9 it follows that the rank of the (m, k)-matrix $(b^{(\kappa)(\alpha)})_{\alpha\kappa}$ is equal to h. The second equality in the same theorem implies then that there are exactly h linearly independent characters among the k induced characters $\Psi^{(\kappa)}$. But $\chi^{(1)}, \chi^{(2)}, \cdots, \chi^{(k)}$ are linear combinations of $\varphi^{(1)}, \varphi^{(2)}, \cdots, \varphi^{(k)}$, and of course vice versa. Hence we have also

THEOREM 10. *The number of linearly independent characters among the k characters of G' induced from the k distinct irreducible characters $\chi^{(\kappa)}$ of G is equal to the number h of those $\mathfrak{K}_\alpha'$ which contain an element of G.*

In particular

COROLLARY.[13] *Let G be an invariant subgroup. The number of the irreducible representations of G mutually non-conjugate in G' is equal to the number h of $\mathfrak{R}'_\alpha$ contained in G.* The former number gives the number of different representations (as well as characters) of G' induced from irreducible representations of G.

This corollary enables us, as one can see without difficulty, to extend Theorem 3 in Nakayama and Shoda, l. c. to the modular case. Namely, the number of distinct irreducible representations of G' by semi-linear transformations in a (not necessarily algebraically closed) field of characteristic p, which correspond to a given isomorphism between a group of automorphisms of the field and the factor group G'/G of G' with respect to the invariant subgroup G, is equal to the above number h, provided that G has k distinct irreducible representations (by linear transformations) in the subfield consisting of all elements invariant under the automorphisms.

4. In this section we consider one group G and its modular representations. Let F_κ, $U_\kappa(\kappa = 1, 2, \cdots, k)$ have the same meaning as in (3). Let, furthermore, $F_1^*, F_2^*, \cdots, F_{k^*}^*$ be distinct irreducible representations of G in an algebraically closed *non-modular* field. Let us consider F_α^* as embedded in the domain of integers for $\mathfrak{p}$ in a suitable algebraic number field Ω, where $\mathfrak{p}$ is a prime ideal divisor of our p in Ω. Take F_α^* mod $\mathfrak{p}$, then we have modular representations $\tilde{F}_\alpha^*$ (in the residue class field with respect to $\mathfrak{p}$). Let $d_{\alpha\kappa}$ denote the multiplicity of F_k in $\tilde{F}_\alpha^*$;

$$\tilde{F}_\alpha^* \leftrightarrow \sum_\kappa d_{\alpha\kappa} F_\kappa .$$

($\leftrightarrow$ expresses that two representations have the same absolutely irreducible constituents. The sum on the right-hand side denotes a representation which may be so reduced that F_κ appear exactly $d_{\alpha\kappa}$ times on the diagonal.) Brauer and Nesbitt called these $d_{\alpha\kappa}$ *the decomposition numbers of G for p.* By means of the decomposition of the determinant $(S(x) + R(y)')$ they proved[14] furthermore the following relation

$$\text{(i)} \qquad c^{(\lambda)(\alpha)} = \sum_{\alpha=1}^{k^*} d_{\alpha\lambda} d_{\alpha\kappa} \qquad (\lambda, \kappa = 1, 2, \cdots, k),$$

which plays a fundamental rôle in their theory. From (i) we have

$$\text{(ii)} \qquad U_\kappa \leftrightarrow \sum_{\alpha=1}^{k^*} d_{\alpha\kappa} \tilde{F}_\alpha^* ,$$

for $U_\kappa \leftrightarrow \sum_\lambda c^{(\lambda)(\kappa)} F_\lambda$.[15]

[13] Cf. Theorem 7 in Nakayama and Shoda, l. c. Cf. also M. Osima, *Beweis eines Satzes in der Darstellungstheorie*, Proc. Imp. Acad. Jap **13** (1937).

[14] They have obtained another proof using Kronecker products.

[15] If we would not wish to use here the fact that the group ring is symmetric, then we should understand U_κ as the indecomposable part of the second regular representation; or, replace $c^{(\lambda)(\kappa)}$ by $\tilde{c}^{(\kappa)(\lambda)}$ understanding U_κ as that of the first regular representation. The consideration below corresponds rather to the latter case.

Conversely, if we assume (ii), then we have certainly (i). Therefore, we may prove (ii) instead of (i).[16] This we do in the following by means of Theorem 8.

For the sake of simplicity (Cf. the lemma below), let us go to the $\mathfrak{p}$-adic extension $\Omega_\mathfrak{p}$ of Ω. Let $\mathfrak{o}$ be the domain of integers in $\Omega_\mathfrak{p}$. Now, let us consider the group ring Γ of G over $\Omega_\mathfrak{p}$. It is semisimple and is a direct sum of k^* simple two-sided ideals Δ_α, which themselves are matric algebras over $\Omega_\mathfrak{p}$; $\Gamma = \Delta_1 + \Delta_2 + \cdots + \Delta_{k^*}$. Let here Δ_α be so enumerated that F_α^* can be defined by a simple left- (or right-) ideal contained in Δ_α. Consider the subring $\mathfrak{J}$ of Γ consisting of all linear combinations of elements of G with coefficients from $\mathfrak{o}$. $\mathfrak{J}$ is an order (with respect to $\mathfrak{o}$) in the sense of arithmetic in algebras. The residue class ring $A = \mathfrak{J}/\mathfrak{p}\mathfrak{J}$ of $\mathfrak{J}$ with respect to the two-sided ideal $\mathfrak{p}\mathfrak{J}$ is just the group ring of G over the modular field $\mathfrak{o}/\mathfrak{p}$. Let us assume here that all F_κ lie already in $\mathfrak{o}/\mathfrak{p}$. Let $A = e_1A + e_2A + \cdots + e_lA$ be a decomposition of A into a direct sum of indecomposable right ideals, where e_i are mutually orthogonal idempotent elements in A with the sum 1 (the unit element of A).

LEMMA. *There exist l mutually orthogonal idempotent elements $\epsilon_1, \epsilon_2, \cdots, \epsilon_l$ in $\mathfrak{J}$ with the sum 1 (the unit element of Γ) such that the residue class of ϵ_i mod $\mathfrak{p}\mathfrak{J}$ is e_i.* Let us assume the lemma for a moment; we shall prove it later.

We consider furthermore a maximal order $\mathfrak{J}'$ of Γ which contains our $\mathfrak{J}$. Consider $\mathfrak{p}\mathfrak{J}' \frown \mathfrak{J}$ and $C = (\mathfrak{p}\mathfrak{J}' \frown \mathfrak{J})/\mathfrak{p}\mathfrak{J}$. C is a nilpotent two-sided ideal of A, hence is contained in the radical of A. Therefore A/C has also k distinct irreducible representations which are essentially equivalent to $F_1, F_2, \cdots, F_k$. Moreover, $A/C = \mathfrak{J}/\mathfrak{p}\mathfrak{J}' \frown \mathfrak{J} \cong (\mathfrak{J}, \mathfrak{p}\mathfrak{J}')/\mathfrak{p}\mathfrak{J}'$ can be considered as a subalgebra of the residue class ring $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ of $\mathfrak{J}'$ with respect to $\mathfrak{p}\mathfrak{J}'$. Take an indecomposable right ideal e_iA of A and let U_κ be the corresponding indecomposable part. The right ideal $(e_iA, C)/C$ of A/C is obviously a direct summand of A/C corresponding to F_κ (in the sense of Theorem 1). Let us construct the right ideal of $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ induced by this right ideal of the subalgebra A/C; we consider namely $(\epsilon_i\mathfrak{J}', \mathfrak{p}\mathfrak{J}')/\mathfrak{p}\mathfrak{J}'$ with the above ϵ_i. Let W denote the representation of G defined by this induced right ideal of $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$. (A left or right ideal of $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ can be considered as a representation module of G with respect to $\mathfrak{o}/\mathfrak{p}$.) Then

$$W \leftrightarrow U_\kappa .$$

To prove this, observe that $(\epsilon_i\mathfrak{J}', \mathfrak{p}\mathfrak{J}')/\mathfrak{p}\mathfrak{J}' \cong \epsilon_i\mathfrak{J}'/\epsilon_i\mathfrak{J}' \frown \mathfrak{p}\mathfrak{J}' = \epsilon_i\mathfrak{J}'/\mathfrak{p}\epsilon_i\mathfrak{J}'$, for since ϵ_i is idempotent $\epsilon_i\mathfrak{J}' \frown \mathfrak{p}\mathfrak{J}' = \mathfrak{p}\epsilon_i\mathfrak{J}'$. Let $(v_1, v_2, \cdots, v_s)$ be a so-called minimal basis of $\epsilon_i\mathfrak{J}'$ with respect to $\mathfrak{o}$. (s is the rank of the right ideal $\epsilon_i\Gamma$ of Γ.) Let W^* be the (non-modular) representation of G defined by this basis (v_ρ) of $\epsilon_i\Gamma$. Then it is easy to see that if we take W^* mod $\mathfrak{p}$ we obtain W. Similarly, let $(u_1, u_2, \cdots, u_s)$ be a minimal basis of $\epsilon_i\mathfrak{J}$, and let U^* be the representation of

---

[16] Here, as well as in the proof below, we do not, contrary to §3, use the fact that the number k of the distinct irreducible representations is equal to the number of the conjugate classes containing elements of an order prime to p; We assume only §§2 and 3 in B.N.M.

G defined by this basis of $\epsilon_i\Gamma$. If we take U^* mod $\mathfrak{p}$ then we get U_κ, as one can see in the same way as above. We have namely derived W and U_κ by taking two equivalent representations mod $\mathfrak{p}$. Hence $W \leftrightarrow U_\kappa$.

On the other hand, the residue class ring $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ is semisimple and is a direct sum of k^* simple two-sided ideals, which themselves are matric algebras over $\mathfrak{o}/\mathfrak{p}$. Let D_α be those two-sided ideals, where they are so enumerated that D_α is represented by $\mathfrak{J}' \frown \Delta_\alpha$. Let L_α be a simple left ideal of $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ contained in D_α. If a basis of L_α (with respect to $\mathfrak{o}/\mathfrak{p}$) is given, then we can choose representatives of those basic elements so that they form a basis of a simple left ideal of Γ contained in Δ_α (with respect to $\Omega\mathfrak{p}$). Therefore, the representation of G defined by L_α is $\tilde{F}^*_\alpha$ (with regard to a suitable embedding of $\tilde{F}^*_\alpha$). A simple right ideal in D_α defines $\tilde{F}^*_\alpha$ too.

Now, Theorem 6, 7 and 8, applied to our $\mathfrak{J}'/\mathfrak{p}\mathfrak{J}'$ and its subring A/C, show then $W \leftrightarrow \sum_\alpha d_{\alpha\kappa}\tilde{F}^*_\alpha$, since $\tilde{F}^*_\alpha \leftrightarrow \sum_\kappa d_{\alpha\kappa}F_\kappa$ by definition. From $W \leftrightarrow U_\kappa$ we have the desired relation (ii).

The only thing left to be proved is the above lemma.

Proof of the Lemma. We first consider a general ring A with the unit element 1 and satisfying the minimum and maximum conditions for left and right ideals. Let N be any nilpotent two-sided ideal of A; N is not necessarily the radical. Let $\bar{c}_1, \bar{c}_2, \cdots, \bar{c}_t$ be mutually orthogonal idempotent elements in the residue class ring A/N with the sum $\bar{1}$ (the unit element of A/N). Then we can prove the existence of t mutually orthogonal idempotent elements c_i in A with the sum 1 such that $c_i \epsilon \bar{c}_i$; the proof runs exactly in the same way as Chap. II, (6), Theorem 1 of Deuring's book referred to in the footnote 3. (The theorem we have used in (1) already.)

Now, coming back to our lemma, let $\omega_1, \omega_2, \cdots, \omega_l$ ($\epsilon\mathfrak{J}$) be any representatives of the residue classes $e_1, e_2, \cdots, e_l$, and apply the above fact first to $A = \mathfrak{J}/\mathfrak{p}^2\mathfrak{J}$ and $N = \mathfrak{p}\mathfrak{J}/\mathfrak{p}^2\mathfrak{J}$. Then we see the existence of l elements ω_i' in $\mathfrak{J}$ such that $\omega_i' \equiv \omega_i \bmod \mathfrak{p}\mathfrak{J}$, $\omega_i'\omega_j' \equiv \begin{cases} \omega_i' \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} \bmod \mathfrak{p}^2\mathfrak{J}$, and $\sum \omega_i' \equiv 1 \bmod \mathfrak{p}^2\mathfrak{J}$. Next we apply the above fact to $A = \mathfrak{J}/\mathfrak{p}^3\mathfrak{J}$ and $N = \mathfrak{p}^2\mathfrak{J}/\mathfrak{p}^3\mathfrak{J}$. Proceeding in this way we get l sequences

$$\omega_i, \omega_i', \omega_i'', \cdots \qquad (i = 1, 2, \cdots, l)$$

in $\mathfrak{J}$ convergent in the sense of $\mathfrak{p}$-adic evaluation. Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_l$ be their limits. Then these ϵ_i have the desired properties.

Added March 29, 1938: Theorem 9 can be proved also by direct calculation of modular characters. By such a method M. Osima (of Osaka) has obtained independently the whole content of (3); he did not publish his results because of the present note.

The Institute for Advanced Study.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

THE UNIQUENESS OF THE POWER SERIES REPRESENTATION OF CERTAIN FIELDS WITH VALUATIONS[1]

BY SAUNDERS MAC LANE

(Received December 27, 1937)

1. Introduction

Some recent investigations of ordered fields, of algebraic functions of several variables,[2] of the arithmetic of general rings,[3] and of the local class field theory,[4] have depended essentially upon the notion of a general valuation of a field K. Such a valuation is a function $Va = \gamma$, defined for a in K and for γ in an ordered Abelian group Γ. This V is assumed to have the same formal properties as in the ordinary case[5] where γ is a real number. Every field with such a general valuation can be embedded, as Krull has shown, in a larger field which is perfect with respect to an extension of the given valuation. In the usual case when Γ is the additive group of real numbers, a perfect extension is one which is topologically complete, and the smallest such perfect extension of K is uniquely determined, up to isomorphism, by the given K. Is this also true for the general valuations? This query, propounded by Krull, will here be considered for the case when the group Γ of values is discrete and of finite rank. We show that the perfect extension is uniquely determined if the characteristic of a certain residue-class field (the next to the last residue-class field in a certain sequence of fields) is 0, and we give examples in §7 to show that the smallest perfect field is not always unique in the other cases.

The formulation of these examples requires a specific construction of perfect extensions. The construction used by Krull[2] for this purpose proceeded indirectly in terms of certain maximal extensions. Our construction given in §§5 and 6 yields a direct existence proof[6] for the immediate perfect extension of a field with a valuation of finite rank. As by-products we obtain a result

---

[1] Presented to the American Mathematical Society, December 28, 1937.

[2] W. Krull, *Allgemeine Bewertungstheorie*, Journal für die Mathematik, 167 (1932), pp. 160–196. Referred to as Krull I.

[3] W. Krull, *Beiträge zur Arithmetik kommutativer Integritätsbereiche*, I, Mathematische Zeitschrift, 41 (1936), pp. 545–577; II, pp. 665–679.

[4] O. F. G. Schilling, *Arithmetic in fields of formal power series in several variables*, Annals of Mathematics, 38 (1937), pp. 551–576.

[5] The definition of a valuation is given in papers cited above, in B. L. van der Waerden, *Moderne Algebra*, vol. I, 1st edition, §65, and in S. Mac Lane, *A construction for absolute values in polynomial rings*, Transactions of the American Mathematical Society, 40 (1936), pp. 363–395.

[6] The possibility of such a simplified construction was suggested to me by O. Ore. The construction, as given in §§5 and 6, can be read independently of §§3 and 4.

on the structure of p-adic fields (Theorem 5 in §5) and a demonstration that in the discrete finite rank case maximal extensions are identical with perfect extensions.

The simplest example of a perfect field is the field of all power series in m variables with coefficients in a given field $\mathfrak{R}$ and with a suitably defined discrete valuation of rank m. Known theorems about the structure of discrete perfect fields[7] show that perfect fields with residue-class fields of characteristic 0 can be represented as power series fields. Hence our fundamental uniqueness theorem for perfect extensions for the case of characteristic 0 states that if a field with a discrete valuation can be isomorphically mapped in a power series field in such a way as to preserve values, then there is, except for automorphisms of the power series field, at most one such map.

2. Perfect fields and their maps

An ordered Abelian group[8] Γ is a set with the properties:

(i) Γ is an additive Abelian group;
(ii) The set Γ is linearly ordered by a relation $\alpha < \beta$;
(iii) If $\gamma < \delta$ in Γ and α is in Γ, then $\gamma + \alpha < \delta + \alpha$.

A subgroup $\Delta \subset \Gamma$ is *isolated* if $0 < \gamma < \delta$ with δ in Δ and γ in Γ imply γ in Δ. If Γ contains but a finite number of isolated subgroups $\Gamma^{(i)}$, they then form a sequence $\Gamma^{(0)} \subset \Gamma^{(1)} \subset \Gamma^{(2)} \subset \cdots \subset \Gamma^{(m-1)} \subset \Gamma$ where $\Gamma^{(0)}$ contains only 0. The number m of isolated subgroups is the *rank* of Γ. If each factor group $\Gamma^{(i)}/\Gamma^{(i-1)}$, $i = 1, \cdots, m$, is isomorphic to the group of rational integers, then Γ is *discrete*.[9] A discrete group Γ of rank m can be generated by m free generators $\alpha_1, \alpha_2, \cdots, \alpha_m$ such that $\alpha_m \gg \alpha_{m-1} \gg \cdots \gg \alpha_1 > 0$, where $\beta \gg \gamma$ means $\beta > k\gamma$ for every positive integer k. Hence any two discrete groups of rank m are isomorphic, preserving order.

A *valuation* $V(K) = \Gamma$ of a field K with a *value group* Γ is a many-one correspondence $a \to Va$ of K^* to Γ, with a in K^*, Va in Γ, such that

$$V(a + b) \geqq \text{Min } (Va, Vb), \qquad V(ab) = Va + Vb \tag{1}$$

holds for all a and b in K^*. Here K^* denotes the set of all elements $a \neq 0$ in K. It is sometimes convenient to define in addition $V0 = \infty$. If Γ and $\bar{\Gamma}$ are two ordered groups, and ϕ an isomorphism of Γ to $\bar{\Gamma}$ which preserves order, then the valuation $\bar{V}$ of K defined by $\bar{V}a = \phi(Va)$ will be said to be *equivalent* to the given V.

A *homomorphism* H of a field K on the set $\{\mathfrak{R}, \infty\}$, which is composed of the

[7] H. Hasse and F. K. Schmidt, *Die Struktur diskret bewerteter Körper*, Journal für die Mathematik, 170 (1934), pp. 4–63. O. Teichmüller, *Diskret bewertete perfekte Körper mit unvollkommenem Restklassenkörper*, Journal für die Mathematik, 176 (1937), pp. 141–152. E. Witt, *Zyklische Körper und Algebren der Charakteristik p vom Grade p^n*, Journal für die Mathematik, 176 (1937), pp. 126–140.

[8] For these definitions, cf. also Krull I, or Schilling, loc. cit.

[9] Krull I, p. 171, §5; Schilling, loc. cit., p. 552.

elements of the field $\mathfrak{K}$ and the symbol ∞, is a many-one correspondence $a \to Ha$ of the set K to the set $\{\mathfrak{K}, \infty\}$ such that (i), if $Ha \neq \infty$ and $Hb \neq \infty$ for any a and b in K, then

$$H(a+b) = Ha + Hb, \qquad H(ab) = (Ha)(Hb) \tag{2}$$

and (ii), for any $a \neq 0$ of K, $Ha = \infty$ if and only if $Ha^{-1} = 0$. If ϕ is an isomorphism of the field $\mathfrak{K}$ to a field $\mathfrak{K}'$, then the homomorphism $H' = \phi H$ is said to be *equivalent* to H.

For a given H, the set K_H of all a in K with $Ha \neq \infty$ and $Ha \neq 0$ forms, by (2), a multiplicative subgroup of the group K^* of all $a \neq 0$ in K. The factor group K^*/K_H, written additively, is called the *value-group* of H. In K^*/K_H denote the coset which contains a given a of K^* by $\alpha = Va$. Two cosets $\alpha = Va$ and $\beta = Vb$ can be ordered by the specification that $\alpha > \beta$ if and only if $H(ab^{-1}) = 0$. This order is independent of the choice of the representatives a and b for α and β, and makes the value-group $\Gamma = K^*/K_H$ an ordered Abelian group. The function Va is a valuation $V(K) = \Gamma$. Any valuation equivalent to this V is called a valuation *associated* with the given homomorphism H.

Conversely, for a valuation $V(K) = \Gamma$ the set P of elements b of positive value ($Vb > 0$) is a prime ideal in the ring R of the elements of K of non-negative value. The field $\mathfrak{K} = R/P$ is the *residue-class field* of V. If a is in R, denote by Ha the residue-class of a in $\mathfrak{K}$, and if a is not in R, set $Ha = \infty$. Then H is a homomorphism of K on $\{\mathfrak{K}, \infty\}$. H and $\mathfrak{K}$ are characterized, up to an isomorphism of $\mathfrak{K}$, by the requirement that $Ha = 0$ holds if and only if $Va > 0$. This H or any equivalent homomorphism is called a homomorphism *associated* with the valuation $V(K) = \Gamma$. The known reciprocal relation of valuations and homomorphisms can be formulated thus:

LEMMA 1. *Every homomorphism $HK = \{\mathfrak{K}, \infty\}$ with a value group $\Gamma = K^*/K_H$ generates an associated valuation $V(K) = \Gamma$ with a residue-class field isomorphic to $\mathfrak{K}$. Conversely, a valuation $V(K) = \Gamma$ with residue-class field $\mathfrak{K}$ generates an associated homomorphism $HK = \{\mathfrak{K}, \infty\}$ with a value group isomorphic to Γ. If V is a valuation associated with H, then H is a homomorphism associated with V, and conversely.*

A valuation V or a homomorphism H has rank m or is discrete according as the corresponding value-group Γ has rank m or is discrete. Henceforth we consider only V, H, or Γ of finite rank. Any rank m valuation can be decomposed into m rank 1 valuations.[10] The associated rank m homomorphism can correspondingly be represented as a product of m rank 1 homomorphisms,[11]

$$H = H' \cdot H'' \cdot \ldots \cdot H^{(m)}, \tag{3}$$

$$H^{(m)}K = \{K^{(m-1)}, \infty\}, \; H^{(m-1)}K^{(m-1)} = \{K^{(m-2)}, \infty\}, \ldots, H'K' = \{\mathfrak{K}, \infty\}. \tag{4}$$

[10] The "Aufspaltungsprozess," Krull I, p. 171 ff.

[11] Schilling, loc. cit., p. 554, gives the decomposition of a homomorphism parallel to Krull's "Aufspaltung."

This decomposition is unique up to isomorphism; that is, if there is any other decomposition $H = G' \cdot G'' \cdot \cdots \cdot G^{(m)}$ into rank 1 homomorphisms $G^{(i)}$, then there is for each i an isomorphism ϕ_i of $K^{(i)}$ onto another field, such that each $G^{(i)}$ is $\phi_{i-1} H^{(i)} \phi_i^{-1}$. Hence each successive residue-class field

$$K^{(m-1)}, \cdots, K'', K', \mathfrak{R} \tag{5}$$

is uniquely determined, except for an isomorphism, by the original H. This can be established by an induction on the rank, using the fact that the decomposition (4) corresponds to a unique sequence[12] of m prime ideals in the valuation ring of V in K.

A homomorphism H of rank 1 is *perfect* if any associated valuation V is perfect; that is, if K is complete with respect to the topology given by the norm $\| a \| = \exp(-Va)$. A homomorphism H of rank m is *step-perfect*[13] if the decomposition of H into rank 1 homomorphisms (4) yields only perfect rank 1 homomorphisms. If H is step-perfect, any associated V is also called step-perfect.[14]

The analytic isomorphism of two valuations can be described thus: If $H_1K_1 = \{\mathfrak{R}, \infty\}$ and $H_2K_2 = \{\mathfrak{R}, \infty\}$ are two homomorphisms on the same field $\mathfrak{R}$, then T is a *map* of H_1 in K_1 on H_2 in K_2 if $a \leftrightarrow Ta$ is an isomorphism of K_1 on K_2 such that $H_1a = H_2Ta$ for all a in K_1. Alternatively, a *map* T of a valuation $V_1(K_1) = \Gamma$ on another valuation $V_2(K_2) = \Gamma$ with the same value group is an isomorphism T of K_1 to K_2 such that $V_2Ta = V_1a$ holds for all a in K_1 (i.e., values are preserved). A map T of V_1 on V_2 is also a map of (suitably chosen) associated homomorphisms H_1 and H_2, and conversely. In the presence of such a map K_1 and K_2 are called *analytically isomorphic*.[15]

A homomorphism $GL = \{\mathfrak{L}, \infty\}$ is called an extension of $HK = \{\mathfrak{R}, \infty\}$ if $\mathfrak{R} \subset \mathfrak{L}$, $K \subset L$, and if $Gb = Hb$ for all b in K. By identifying each coset of K^*/K_H with the unique coset of L^*/L_G in which it is contained, the value group of H can (and will) be considered as a subgroup of the value-group of G. If with this convention the extended homomorphism G has the same value group and the same residue-class field $\mathfrak{R}$ as H, then G is called an *immediate extension* of H.

---

[12] Krull I, §5.

[13] Schilling, loc. cit., p. 554.

[14] Because the rank of Γ is finite, this definition is readily shown equivalent to that given in Krull I, p. 177.

[15] Cf. A. Ostrowski, *Untersuchungen zur arithmetischen Theorie der Körper*, Mathematische Zeitschrift, 39 (1934), pp. 269–404. Our definition of an analytic isomorphism differs only in form from the apparently more inclusive definition of Hasse and Schmidt (loc. cit., p. 27). There T is called an analytic isomorphism of $V_1(K_1) = \Gamma_1$ on $V_2(K_2) = \Gamma_2$ if T is an isomorphism of K_1 on K_2 such that for any a and b in K_1, $V_1a \leqq V_1b$ holds if and only if $V_2Ta \leqq V_2Tb$. For if T has this property, then $V_1a = V_1b$ if and only if $V_2Ta = V_2Tb$, so the correspondence $V_1a \leftrightarrow V_2Ta$ is a one-one order-preserving isomorphism of Γ_1 on Γ_2. If each element of Γ_2 be replaced by the corresponding element from Γ_1, then $V_2Ta = V_1a$, exactly as in our condition (5).

3. Maps on power series fields

If a discrete perfect field K with a rank 1 valuation has a residue-class field $\mathfrak{K}$ with the same characteristic as K, then K is by known structure theorems[16] analytically isomorphic to a power series field. The different possible maps of K on the power series field will now be described by characterizing those subfields of K which can map into the field of coefficients of the power series.

$\mathfrak{K}(\text{series } u)$ will denote the field composed of all formal power series $y = \sum_m^\infty b_i u^i$ with coefficients b_i in a given field $\mathfrak{K}$. The field has a discrete rank 1 series valuation $Wy = m$, for y with first coefficient $b_m \neq 0$, and an associated homomorphism S on $\mathfrak{K}$, with $Sy = \infty$, b_0 or 0 according as $Wy < 0$, $Wy = 0$, or $Wy > 0$. We call S the "series homomorphism" of the field.

THEOREM 1. *Let $HK = \{\mathfrak{K}, \infty\}$ be a discrete rank 1 homomorphism with an associated valuation V, for which the value group consists of all rational integers. If π in K has $V\pi = 1$, and if $M \subset K$ is a subfield with $H(M^*) = \mathfrak{K}^*$, then there exists one and only one isomorphism T of K on the series field $\mathfrak{K}(\text{series } u)$ such that*

$$T\pi = u, \qquad T(M) = \mathfrak{K}, \qquad Ha = STa \text{ for } a \text{ in } K, \tag{6}$$

where S is the series homomorphism of $\mathfrak{K}(\text{series } u)$. Conversely, any analytic isomorphism T of K on $\mathfrak{K}(\text{series } u)$ determines $\pi = T^{-1}u$ and $M = T^{-1}(\mathfrak{K})$ with $V\pi = 1$ and $H(M^*) = K^*$.

PROOF. The construction of a T follows known methods.[17] Were there a second map T', then for any b in M, $T'b = ST'b = Hb = STb = Tb$, so that $Tc = T'c$ for all c in $M(\pi)$, hence for all c in K. The converse is trivial.

A T can thus be constructed if the corresponding "coefficient field" M is found, as for instance by the following embedding process.

LEMMA 2. *If $HK = \{\mathfrak{K}, \infty\}$ is a discrete perfect homomorphism of rank 1, if $N \subset K$ is a subfield such that $H(N^*) \subset \mathfrak{K}^*$, and if $\mathfrak{K}$ is separable over $H(N)$, then there exists a subfield M of K with $N \subset M$ and $H(M^*) = \mathfrak{K}^*$.*

PROOF. Set $\mathfrak{N} = H(N)$. The hypothesis that $\mathfrak{K}:\mathfrak{N}$ is separable means[18] that there exists a well ordering a_λ of all elements of $\mathfrak{N}$ (λ any ordinal number $\lambda < \nu$, ν fixed) such that if $\mathfrak{K}_\lambda$ is the subfield of $\mathfrak{K}$ obtained by adjoining to $\mathfrak{N}$ all elements a_μ preceding a_λ, then a_λ is either transcendental or separable algebraic over $\mathfrak{K}_\lambda$. For this well-ordering we construct, for each λ, a subfield $N_\lambda \subset K$ such that (i) $N_1 = N$; (ii) $H(N_\lambda^*) = \mathfrak{K}_\lambda^*$; (iii) if μ precedes λ, then $N_\mu \subset N_\lambda$. Since the construction of N_1 is given, we can apply transfinite induction and assume N_μ constructed for every $\mu < \lambda$. If λ is a limit-number, let N_λ be the join of all fields N_μ with $\mu < \lambda$. The properties (i) to (iii) follow. If λ has a predecessor, then $N_{\lambda-1}$ is given, and we consider two different cases.

---

[16] Hasse and Schmidt, loc. cit., Theorem A, p. 9; Teichmüller, loc. cit., p. 148; Witt, loc. cit., Satz 8, p. 133.

[17] The proof is substantially that given in terms of valuations by Teichmüller, loc. cit., Hilfsatz 3, p. 143. Cf. also p. 148.

[18] Hasse and Schmidt, loc. cit., p. 15.

CASE 1. $a_{\lambda-1}$ is algebraic over $\mathfrak{R}_{\lambda-1}$. Because of the separability, $a_{\lambda-1}$ is a simple root of a polynomial $f(t) = \sum c_i t^i$ with coefficients in $\mathfrak{R}_{\lambda-1}$. For each c_i, we can by (ii) pick a b_i in $N_{\lambda-1}$ with $Hb_i = c_i$. Construct the polynomial $g(t) = \sum b_i t^i$. Because $a_{\lambda-1}$ was a simple root and because K is perfect with respect to any rank 1 valuation associated with H, the reducibility theorem of Hensel-Rychlik[19] applies and asserts that there is in K a root θ of $g(t) = 0$ with $H\theta = a_{\lambda-1}$. The field $N_\lambda = N_{\lambda-1}(\theta)$ then satisfies our conditions.

CASE 2. $a_{\lambda-1}$ is transcendental over $K_{\lambda-1}$. Then if x is any element of K with $Hx = a_{\lambda-1}$ it follows that x must likewise be transcendental over $N_{\lambda-1}$ and that the field $N_\lambda = N_{\lambda-1}(x)$ satisfies the conditions.

The required field M is the final field N_ν so obtained.

4. The uniqueness of step-perfect extensions

The uniqueness of an immediate step-perfect extension of a given discrete valuation for the case when the next to the last residue-class field has characteristic 0 will now be established by induction on the rank. For this induction we need the following more detailed formulation.

THEOREM 2. *Let $H_i(L_i) = \{\mathfrak{L}, \infty\}$, for $i = 1, 2$ be two step-perfect rank m homomorphisms, with subfields $K_i \subset L_i$ having a common residue-class field $H_i(K_i) = \mathfrak{K}$, for $i = 1, 2$. Let L_i' be the next to the last residue-class field of H_i in L_i. Suppose that K_i and L_i have the same value group under H_i, for $i = 1, 2$, and that either*

(i) *$\mathfrak{L}$ has characteristic 0, or*
(ii) *$\mathfrak{K} = \mathfrak{L}$, and L_1' has characteristic 0.*

Then any analytic isomorphism T of K_1 to K_2 (i.e., $H_2Ta = H_1a$) can be extended to an analytic isomorphism τ of L_1 to L_2.

PROOF. Note that L_i' is the field arising in the decomposition

$$(7) \qquad H_i = H_i' \cdot G_i, \qquad G_i(L_i) = \{L_i', \infty\}, \qquad H_i'(L_i') = \{\mathfrak{L}, \infty\}$$

of H_i into a product of two homomorphisms G_i and H_i' of respective ranks $m - 1$ and 1.

CASE 1. $m = 1$, condition (i) obtains, and $H_1(K_1) = \{\mathfrak{K}, \infty\}$ is perfect. The required τ can then be constructed by the power series representation. Because of (i), the prime subfield R_0 of K_1 is of characteristic 0 and has $H_1(R_0^*) \subset \mathfrak{K}^*$. By Lemma 2, R_0 can then be embedded in a "coefficient field" $M_1 \subset K_1$ with $H_1(M_1^*) = \mathfrak{K}^*$. Then $T(M_1) = M_2$ has $H_2(M_2^*) = \mathfrak{K}^*$, and by Lemma 2 there also exist "coefficient fields" N_i for the extended fields L_i, such that $M_i \subset N_i \subset L_i$ and $H_i(N_i^*) = \mathfrak{L}$, where $i = 1, 2$. Let V_i be a valuation associated with H_i. Since H_i is discrete, we can suppose that the value group of V_i is the group of rational integers. Choose π_1 in K_1 with $V_1\pi_1 = 1$. Then $T\pi_1 = \pi_2$ is in K_2 with $V_2\pi_2 = 1$.

[19] Hasse and Schmidt, loc. cit., p. 31, first theorem; K. Rychlik, *Zur Bewertungstheorie der algebraischen Körper*, Journal für die Mathematik, 153 (1924), p. 94; cf. §11.

Now construct the series field $\mathfrak{L}(\text{series } u)$ with a rank 1 series homomorphism S. By the series representation theorem there exist maps T_i of L_i on $\mathfrak{L}(\text{series } u)$ such that $H_i = S \cdot T_i$, $T_i\pi_i = u$ and $T_i(N_i) = \mathfrak{L}$, for $i = 1, 2$. The combination $\tau = T_2^{-1} \cdot T_1$ is then an isomorphism of L_1 on L_2 with $H_1 = H_2 \cdot \tau$.

Finally, τ is an extension of the given T. For if a_1 is in M_1, then $a_2 = Ta_1$ is in M_2, and by (6) in Theorem 1, $T_1a_1 = H_1a_1$ and $T_2a_2 = H_2a_2$. But $H_2a_2 = H_2Ta_1 = H_1a_1$ by hypothesis, so $T_1a_1 = T_2a_2$, and by definition of τ, $\tau a_1 = T_2^{-1}T_1a_1 = a_2$. Hence $\tau a_1 = Ta_1$ for a_1 in M_1. By construction $\tau\pi_1 = T_2^{-1}T_1\pi_1 = T_2^{-1}u = \pi_2 = T\pi_1$; hence $\tau = T$ holds in the field $M_1(\pi_1)$. But M_1 is a "coefficient field" for K_1, so every residue-class of K_1 contains an element of M_1, and $M_1(\pi_1)$ is everywhere dense in K_1. Both τ and T are analytic isomorphisms, hence are continuous in the topologies given by V_1 and V_2, and $\tau = T$ on $M_1(\pi_1)$. Hence $\tau = T$ on all K_1, as asserted.

Case 2. $m = 1$, condition (i) holds, but K_1 is not perfect in the valuation V_1 associated with H_1. Then K_i has a perfect extension $\bar{K}_i \subset L_i$. By the uniqueness of the ordinary perfect extension[20] there exists an extension $\overline{T}$ of T which is a map of $\bar{K}_1$ on $\bar{K}_2$. The previous case then applies.

Case 3. $m = 1$, and condition (ii) holds. The perfect extension $\bar{K}_1$ of K_1 within L_1 has by (ii) the same residue-class field and the same value group as does L_1. Hence $\bar{K}_1 = L_1$. The extension $\overline{T}$ of T to $\bar{K}_1$ is found as before, and is itself the desired map τ.

Case 4. $m > 1$. To make an induction on m, decompose each homomorphism H_i as in (7), thus decomposing the subfield homomorphisms $H_i(K_i) = \mathfrak{R}$ into

$$(8) \qquad H_i = H_i' \cdot G_i, \qquad G_i(K_i) = \{K_i', \infty\}, \qquad H_i'K_i' = \{\mathfrak{R}, \infty\}, \qquad (i = 1, 2).$$

Since such a decomposition is unique up to isomorphism, and since T is an isomorphism of K_1 to K_2 carrying H_1 into H_2, these two decompositions (8) must be isomorphic. Specifically; if we set $T'b = G_2Ta$ for any $b = G_1a$ in K_1', then T' is an isomorphism of K_1' to K_2' with $H_2'T' = H_1'$. As the hypotheses of the theorem apply to these rank 1 homomorphisms H_1', there is an extension τ' of T' which is an isomorphism of L_1' to L_2' with $H_2'\tau'b = H_1'b$ for all b in L_1'. Consider next the rank $m - 1$ homomorphisms

$$(\tau' \cdot G_1)(L_1) = \{L_2', \infty\}, \qquad G_2(L_2) = \{L_2', \infty\}.$$

Under $\tau'G_1$ and G_2 the subfields K_1 and K_2 both have the same map K_2', for by (8)

$$(\tau'G_1)K_1 = \tau'(K_1') = T'(K_1') = \{K_2', \infty\}, \qquad G_2(K_2) = \{K_2', \infty\}.$$

The common residue-class field L_2' has characteristic 0 in either case (i) or (ii). Also, T maps $\tau'G_1$ into G_2, for if a is in K_1, then by definition of T', $G_2(Ta) = T' \cdot G_1a = (\tau'G_1)a$. The induction assumption then applies to the homomor-

[20] See for instance the formulation given by Hasse and Schmidt, loc. cit., p. 29.

phisms $\tau' \cdot G_1$ and G_2, and yields an extension τ of T which is a homomorphism of L_1 to L_2, with $G_2\tau = \tau' G_1$. Hence $H_2\tau = H_2' \cdot G_2\tau = H_2'\tau' G_1 = H_1' G_1 = H_1$, so τ has all required properties.

The uniqueness of immediate step-perfect extensions for suitable residue-class field characteristics now follows at once.

THEOREM 3. *Let $HK = \{\Re, \infty\}$ be a discrete rank m homomorphism with a "next to the last" residue-class field K' of characteristic* 0 (cf. §2, (5)), *and let $F_i(L_i) = \{\Re, \infty\}$ with $L_i \supset K$ for $i = 1, 2$ be two immediate step-perfect extensions of H. Then there exists an isomorphism T of L_1 to L_2 which is the identity on the common subfield K, and which has $F_2 Ta = F_1 a$ for all a in L_1.*

The parallel theorem for valuations is

THEOREM 3A. *Let $V(K) = \Gamma$ be a discrete valuation of rank m for which the next to the last residue-class field has characteristic* 0. *Let $V_1(L_1) = \Gamma$, $V_2(L_2) = \Gamma$ be two immediate step-perfect extensions of $V(K) = \Gamma$. Then there exists an analytic isomorphism T of $V_1(L_1) = \Gamma$ to $V_2(L_2) = \Gamma$, which is the identity on K and which has $F_2 Ta = F_1 a$ for all a in L_1, where each F_i is a homomorphism associated with L_i, $i = 1, 2$.*

5. Extensions of residue-class fields

Our construction method for step-perfect extensions depends essentially on a process for extending the residue-class field of a valuation in any prescribed manner.

THEOREM 4. *If $H(K) = \{\Re, \infty\}$ is any homomorphism and if $\bar{\Re} \supset \Re$ is any extension of the residue-class field, then there is a field $\bar{K} \supset K$ and an extension of H to a homomorphism $\bar{H}(\bar{K}) = \{\bar{\Re}, \infty\}$, such that the value group $\bar{\Gamma}$ of $\bar{H}$ is the same as the value group Γ of H.*

The conclusion $\Gamma = \bar{\Gamma}$ means that the group $\bar{K}^*/\bar{K}_{\bar{H}}$, considered as an extension of K^*/K_H, is actually equal to K^*/K_H. Because of the definition of the value group, this in turn means that for every $b \neq 0$ in $\bar{K}$ there is a $c \neq 0$ in K with the same value; that is, such that $H(bc^{-1})$ is in $\bar{\Re}^*$.

PROOF. CASE 1. $\bar{\Re}$ is a simple transcendental extension $\bar{\Re} = \Re(t)$. Construct a transcendental extension $M = K(x)$ of K. For any polynomial $f(x) = \sum a_i x^i$ in $K[x]$ define

$$\bar{H}(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0) = (Ha_n)t^n + (Ha_{n-1})t^{n-1} + \cdots + (Ha_1)t + (Ha_0), \tag{9}$$

where the right side is to be replaced by ∞ if any coefficient Ha_i is ∞. For a rational function define $\bar{H}(f(x)/g(x)) = (\bar{H}f(x))/(\bar{H}g(x))$, where by suitable normalizations we can assume that $\bar{H}g \neq \infty$; $\bar{H}g \neq 0$. The so defined function[21] $\bar{H}$ is a homomorphism $\bar{H}(K(x)) = \{\Re(t), \infty\}$. To show $\bar{\Gamma} = \Gamma$, let V

[21] The corresponding construction for a valuation is a special case of the "inductive" values discussed in Mac Lane, loc. cit., Theorem 10.2. This special case is also used by W. Krull, *Beiträge* ··· I, loc. cit., p. 559.

be a valuation associated with H and in the polynomial $f(x)$ of (9) pick a j with $Va_j = \text{Min}\,[Va_n, \cdots, Va_0]$. Then $a_j^{-1}f(x)$ has one coefficient 1 and hence $\bar{H}(a_j^{-1}f(x))$ is in $\bar{\mathfrak{R}}^*$. Similarly, for any rational function $f(x)/g(x)$ there is a constant b_k/a_j such that $\bar{H}[(b_k/a_j)(f(x)/g(x))]$ is in $\bar{\mathfrak{R}}^*$. Hence $\bar{\Gamma} = \Gamma$.

CASE 2. $\bar{\mathfrak{R}}$ is a simple algebraic extension $\bar{\mathfrak{R}} = \mathfrak{R}(\theta')$, where θ' is a root of a polynomial $h(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_0$, irreducible in $\mathfrak{R}[z]$. Then choose b_i in K with $Hb_i = c_i$, for $i = 0, \cdots, n-1$ and set

$$g(z) = z^n + b_{n-1}z^{n-1} + \cdots + b_0 .$$

Then, by a form of the general Eisenstein irreducibility criterion, $g(z)$ is irreducible in $K[z]$. Construct $M = K(\theta)$, where θ is a root of $g(z)$, and define a homomorphism[22] of M by

$$\bar{H}(a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1}) = (Ha_0) + (Ha_1)\theta' + \cdots + (Ha_{n-1})\theta'^{n-1}.$$

This homomorphism, $\bar{H}(K(\theta)) = \{\mathfrak{R}(\theta'), \infty\}$ again has the original value group Γ as value group.

CASE 3. $\bar{\mathfrak{R}}$ is any extension whatever of $\mathfrak{R}$. A simple well-ordering argument will reduce the theorem in this case to the two preceding cases.

An application may be made to study the structure of perfect fields of rank 1. A *p-adic valuation* $V(K) = \Gamma$ with the residue-class field $\mathfrak{R}$ is a perfect valuation in which Γ is the group of all rational integers, K and $\mathfrak{R}$ have the characteristics 0 and p respectively, and $Vp = 1$, so that p is an element of minimum positive value in K. Known structure theorems[23] assert that the residue-class field $\mathfrak{R}$ is a characteristic invariant of such a p-adic valuation; that is, two fields with p-adic valuations are analytically isomorphic if and only if they have isomorphic residue-class fields. When one residue-class field is an extension of the residue-class field of another p-adic valuation, we have the following theorem:

THEOREM 5. *If V_i is a p-adic valuation of K_i with a residue-class field $\mathfrak{R}_i$, $i = 1, 2$, and if τ is an isomorphism of $\mathfrak{R}_1$ into a subfield of $\mathfrak{R}_2$, then there is an analytic isomorphism T of K_1 into a subfield of K_2, such that $\tau H_1 a = H_2 Ta$ for any a with $V_1 a \geqq 0$ in K_1, where H_i denotes the residue-class homomorphism $H_iK_i = \{\mathfrak{R}_i, \infty\}$ associated with V_i, $i = 1, 2$.*

PROOF. Since τH_1 is a homomorphism of K_1 on part of $\mathfrak{R}_2$, there is by Theorem 4 an extension $\bar{K}_1 \supset K_1$ and a homomorphism $\bar{H}\bar{K}_1 = \{\mathfrak{R}_2, \infty\}$ which is an extension of τH_1. This $\bar{H}$ is of rank 1, hence K_1 can be embedded in a field M_1, perfect with respect to an extension of $\bar{H}$. This extension M_1 has the same residue-class field $\mathfrak{R}_2$ as does the given K_2; hence M_1 is analytically isomorphic to K_2 by the structure theorem quoted above.[23] This isomorphism

---

[22] The associated valuation is a simple infinite "approximant" of the form $[V, Vz = 0, Vg(z) = \infty]$, in the sense of S. Mac Lane, *A construction for prime ideals as absolute values of an algebraic field*, Duke Mathematical Journal, 2 (1936), pp. 492–510.

[23] Hasse and Schmidt, Theorem B_1, p. 12; Witt, Theorem 10, p. 134; Teichmüller, p. 149.

yields the required map T of the subfield $K_1 \subset M_1$ into K_2. The condition ${}_\tau H_1 a = H_2 T a$, which asserts that T generates the originally given isomorphism τ of the residue-class fields, then follows readily.

Theorem 5 can also be established by the direct but laborious process of repeating the arguments of the Hasse-Schmidt structure theory for this case.

6. Construction of step-perfect extensions

A valuation $W(L) = \Delta$ is an *extension* of a valuation $V(K) = \Gamma$ if $L \supset K$, $\Delta \supset \Gamma$ and $Wa = Va$ for a in K. The extension is *immediate* if $\Delta = \Gamma$ and if the residue-class field of L equals that of K, provided each residue-class of K be identified with the unique residue-class of L in which it is contained.

THEOREM 6. *Every rank m valuation has an immediate extension which is step-perfect.*

PROOF. In terms of the associated homomorphisms, we have to show that any rank m homomorphism $HK = \{\mathfrak{K}, \infty\}$ can be extended to a step-perfect homomorphism $FL = \{\mathfrak{K}, \infty\}$ without extension of the value group. For rank $m = 1$ the procedure is well known,[24] hence we make the induction assumption that the theorem is true for rank $m - 1 > 0$. Decompose H into the product

$$H = H' \cdot G, \qquad GK = \{K', \infty\}, \qquad H'K' = \{\mathfrak{K}, \infty\},$$

where G has rank $m - 1$, H' rank 1. By the rank 1 case, there is an immediate perfect extension $\bar{H}'\bar{K}' = \{\mathfrak{K}, \infty\}$ of H'. By Theorem 4, there is then an extension $\bar{G}\bar{K} = \{\bar{K}', \infty\}$ with the same value group as G, and with extended residue-class field $\bar{K}' \supset K'$. By the induction assumption, there is an immediate step-perfect extension $\bar{G}_1(L) = \{\bar{K}', \infty\}$ of $\bar{G}$. The product

$$\bar{H} = \bar{H}'\bar{G}_1, \qquad \bar{G}_1(L) = \{\bar{K}', \infty\}, \qquad \bar{H}'(\bar{K}') = \{\mathfrak{K}, \infty\}$$

is then step-perfect by definition, since both of its factors are step-perfect. The value group of $\bar{H}$ is that of H, because the analogous fact holds for the two factors. Specifically, (cf. discussion following Theorem 4) if $b \neq 0$ in L, there is in K an element $c \neq 0$ with the same value, so that $d = \bar{G}_1(bc^{-1})$ is in $\bar{K}'^*$. To d there is in turn an $f = Ge$ in K'^* with the same value in the valuation associated with $\bar{H}'$, so that $\bar{H}'(df^{-1})$ is in $\mathfrak{K}^*$. Then

$$\bar{H}(b(ce)^{-1}) = \bar{H}'[\bar{G}_1(bc^{-1})\bar{G}_1 e^{-1}] = \bar{H}'[df^{-1}] \in \mathfrak{K}^*.$$

Thus b and ce have the same value in the final valuation, so that H and $\bar{H}$ have identical value groups.

A maximally perfect field K is one with a valuation $V(K) = \Gamma$ which has no proper immediate extension. Hence Theorem 6 gives immediately a proof[25] of

---

[24] Hasse and Schmidt, p. 24.

[25] Krull, I, §14, gives a longer proof for the case of arbitrary rank.

THEOREM 7. *Every maximally perfect valuation of finite rank is step-perfect.*

Conversely, in the discrete case a simple induction yields the result:

THEOREM 8. *A discrete valuation of finite rank is step-perfect if and only if it is maximally perfect.*

Consequently, in investigations of discrete finite-rank valuations, the notion of maximally perfect can be replaced by the more explicit notion of step-perfection.

7. Examples of non-unique extensions

The sequence of fields $K = K^{(m)}, K^{(m-1)}, K^{(m-2)}, \cdots, K', \mathfrak{R}$ for a rank m valuation of K is always such that all the fields in this sequence have the same characteristic or such that the first s fields ($0 < s \leqq m$) have characteristic 0 and all the remaining fields have a fixed characteristic p. The uniqueness theorem for immediate step-perfect extensions was proven only in the case when at most the last field has characteristic p. In all the remaining cases the immediate step-perfect extension of K is not always uniquely determined, as we now indicate for the case $m = 2$ by two examples for which the sequence of residue-class field characteristics are (p, p, p) and $(0, p, p)$ respectively.

EXAMPLE I. Let P the Galois field with p elements; let S_1 be the series homomorphism of P(series x) on P, and let S_2 be the series homomorphism of $K = P$(series x, series y) on P(series x). Here K is the field of power series in y with coefficients themselves series in x, and the product $S = S_2 \cdot S_1$ is a step-perfect[26] rank 2 homomorphism of K on P. Denote by V an associated valuation. By cardinal number considerations there is in P(series x) an element z with $Vz = 0$ which is transcendental over $P(x)$. Then $K_0 = P(x, z^p, y)$ is a subfield of K, which under S has the same value group and the same residue-class field as does K. Thus K is an immediate step-perfect extension of K_0.

Construct an automorphism T of K_0 by specifying

$$(10)\quad Tx = x, \qquad Tz^p = z^p + y, \qquad Ty = y, \qquad Ta = a \text{ for } a \text{ in } P.$$

That T is an automorphism follows because the elements x, z^p, y are, like the elements x, $z^p + y$, y, three independent indeterminates over P. This T will be analytic if either one of the conditions

$$(11)\qquad STb = Sb, \qquad\qquad VTb = Vb$$

holds for all b in K_0.

To prove (11) consider any

$$b = b_0 + b_1 z^p + \cdots + b_n (z^p)^n \qquad \text{in } P(x)[z^p].$$

Then by (10)

$$Tb = b_0 + b_1(z^p + y) + \cdots + b_n(z^p + y)^n = b + y \cdot g(z^p, y)$$

[26] Schilling, loc. cit., p. 556, Lemma 1.

where $g(z^p, y)$ is a polynomial with coefficients in $P(x)$. Hence, by the definition of S, (11_1) and therefore also (11_2) holds for b in $P(x)[z^p]$. If next $b = f(z^p, y)$ is a polynomial with coefficients in $P(x)$, and if b is written as a polynomial in y, then Vb is the value of the constant term. Hence by a simple expansion of Tb, $VTb = Vb$. The same fact follows for a rational function b by use of the elementary properties of V. Hence T is an analytic automorphism of K_0.

If K_0 had a uniquely determined step-perfect extension, in the sense formulated in Theorem 3A, it would follow that T could be extended to be a map $\overline{T}$ of the immediate step-perfect extension $K \supset K_0$ onto the step-perfect extension $K \supset TK_0$. Then $(\overline{T}z)^p = \overline{T}z^p = z^p + y$, involves the first power of y. This is impossible, since any series $\overline{T}z$ in y in a field of characteristic p has a p-th power involving only terms in y^p.

The inner reason why the step-perfect extension of K_0 is not unique is the fact that the rank 1 homomorphism S_2 has for K and K_0 the residue-class fields $P(\text{series } x)$ and $P(x, z^p)$ respectively. The former is not a separable infinite extension of the latter in the sense used in the proof of Lemma 2, and thus in our uniqueness proof.

EXAMPLE II. Sequence of characteristics, $(0, p, p)$. By the method of Theorem 4 a rank 2 step-perfect extension of the p-adic numbers will be constructed in two essentially different ways. As in Example I, take the element z in $P(\text{series } x)$ and the series homomorphism S' of $P(\text{series } x)$ on P. Let K now represent the field of p-adic numbers, W the p-adic valuation of K, and G the associated homomorphism of K on P. Consider the field $L = K(s, r)$ obtained from K by adjoining two indeterminates s and r. As in the first case of Theorem 4, extend G to a homomorphism

$$(12) \qquad G(L) = \{P(x, z^p), \infty\}, \qquad Gs = x, \qquad Gr = z^p.$$

Then set $r_1 = r$, $r_2 = r(1 + p)$; so that $Gr_1 = Gr_2 = z^p$.

The polynomials $y^p - r_1$ and $y^p - r_2$ are both irreducible in $L[y]$. For were either polynomial reducible it would, by the Gauss lemma, be reducible in $R[y]$, where R is the valuation ring composed of all elements b of L with $Gb \neq \infty$. Hence were either polynomial reducible, its image $y^p - Gr_i = y^p - z^p$ would be reducible in $G(R) = P(x, z^p)$. But $y^p - z^p$ is reducible in a field of characteristic p only if it decomposes into linear factors $(y - z)^p$, which is impossible since z is not in the two-fold transcendental field $P(x, z^p)$.

Construct the fields $L_i = L((r_i)^{1/p})$, and, as in the second case of Theorem 4, extend the homomorphism G of L to two homomorphisms

$$G_i(L_i) = \{P(x, z), \infty\}, \qquad G_i r_i = z, \qquad i = 1, 2.$$

Theorem 4 then gives a further extension of G_i to a rank one perfect homomorphism

$$(13) \qquad G_i(M_i) = \{P(\text{series } x), \infty\} \; M_i \supset L_i, \qquad i = 1, 2,$$

which has the same value-group as the original G. Using the series homomorphism S' each product $H_i = S'G_i$ is a rank 2 homomorphism of the corresponding field M_i onto P. H_i is step-perfect, for both factors S' and G_i are perfect. Thus H_1 and H_2 are two step-perfect immediate extensions of the rank 2 homomorphisms $H = S'G$ of L on P. If the immediate step-perfect extension of H were unique there would be an analytic isomorphism T with

$$T(M_1) = M_2, \qquad H_2Ta = H_1a \text{ for } a \text{ in } M_1, \qquad Tb = b \text{ for } b \text{ in } L.$$

M_2 then contains an element $u = r_2^{p^{-1}}/Tr_2^{p^{-1}}$. By the definitions of r_1 and r_2,

$$u^p = r_2/Tr_1 = r(1 + p)/Tr = r(1 + p)/r = 1 + p.$$

Therefore, using the binomial expansion of $u^p = [(u - 1) + 1]^p$, we have

$$(14) \quad 0 = (u - 1)^p + p(u - 1)^{p-1} + \binom{p}{2}(u - 1)^{p-2} + \cdots + p(u - 1) + p.$$

Let $\overline{W}$ be a valuation associated with the rank 1 homomorphism G_2 of (13). As G_2 is an extension of the homomorphism G of the p-adic field K, $\overline{W}$ is an immediate extension of the p-adic valuation W of K in which Wp is the minimum positive value. Therefore $\overline{W}p$ is the minimum positive value in the value-group of $\overline{W}$. But in (14) each binomial coefficient has value at least Wp. Since at least two terms of (14) must have the same minimum value (a Newton polygon argument), it follows that $\overline{W}(u - 1) = 1/p\ \overline{W}p$, which is a positive value smaller than the minimum positive value, a contradiction. The assumed uniqueness of a step-perfect extension of L is thus false.

THE UNIVERSITY OF CHICAGO.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ON CHAINS OF TOPOLOGICAL SPACES[1]

BY S. LEFSCHETZ

(Received June 20, 1937, revised Feb. 9, 1938)

INTRODUCTION

The object of the present paper is to utilize the general Vietoris scheme to set up a new system of chains for compact metric spaces. The new chains which we call $\mathfrak{P}$-chains (abridged from projection-chains) obey the axioms of DJ1 (No. 1), except axiom Vd, the dimensional part of axiom V, which is too stringent and not needed in practice. The $\mathfrak{P}$-chains are more satisfactory in every way than the projection-chains of Am. J. and DJ1 (p. 4). In particular they have topological character, being in substance transformed into similar chains when the space merely undergoes a continuous single-valued transformation. Generally speaking they are as plastic as singular chains. However we experienced a number of difficulties in the treatment of the chain groups, and in this part particularly (and in other parts also) we derived much assistance from Dr. N. E. Steenrod. The $\mathfrak{P}$-chains give rise to a theory of local connectedness in the sense of homology, which we have developed in full, although only a few indications will be given here. This theory culminates in an extension of our fixed point theorem to which the extremely simple proof which we gave recently in Ann. M. 2 is also applicable.

Here as in Ann. M. 2 we lean ever more heavily upon the general theory of abstract complexes. For this reason we shall first consider somewhat fully (§I) certain results which are borrowed from their theory. Theorem 1, on chain-deformations, which we prove in §I, has considerable interest for its own sake, and many other applications than those given here.

I. ABSTRACT COMPLEXES AND THEIR CHAIN-DEFORMATIONS

1. The abstract elements: complex, chain, cycle, simplicial or otherwise, and their products shall be understood in the sense of A. W. Tucker[2] or of Bull MS.

[1] We shall have occasion to use the following references. After each reference we give in square brackets its designation through the rest of the paper:

S. Lefschetz: Colloquium Lectures, *Topology*, New York (1930) [Topology].
Annals of Mathematics, vol. 35 (1934), pp. 118–129 [Ann. M. 1]
Duke Journal, vol. 1, (1935), pp. 1–19 [DJ1]
Duke Journal, vol. 2 (1936), pp. 435–442 [DJ2]
American Journal of Mathematics, vol. 55 (1933), pp. 469–504 [Am. J.]
Bulletin American Mathematical Society, vol. 43 (1937), pp. 345–359 [Bull. MS].
Annals of Mathematics, vol. 38 (1937) pp. 819–822, [Ann. M. 2]

[2] A. W. Tucker, Annals of Mathematics, vol. 34 (1933) pp. 191–243.

Until we come to the $\mathfrak{P}$-chains (§II), the group of coefficients is the additive group of all integers. The allowable chains are of dimensions $q = -1, 0, 1, \cdots$, those of dimension -1 being merely integers (with a general additive abelian group of coefficients $\mathfrak{G}$ they would be elements of $\mathfrak{G}$). The boundary of an $F(C_0)$ is thus a C_{-1}, i.e. an integer (with the general group $\mathfrak{G}$ an element of $\mathfrak{G}$). By definition $F(C_{-1}) = 0$, and hence for all admissible chains $FF = 0$. When the chains are simplicial $F(C_0)$ is the sum of the coefficients of the points entering in C_0, that is to say the Kronecker-index of C_0, usually designated by (C_0). Thus C_0 is only a cycle when $(C_0) = 0$. This is the most convenient scheme for local connectedness, and corresponds to: a point A is not a cycle, the difference $A - B$ of two points is a cycle.

Noteworthy products are $\sigma_1 \times C_p$ of a one-simplex by a p-chain. Here C_p may be a chain of any type. However if C_p is simplicial the product-cells $\sigma_1 \times \sigma$ are "prismatic." In a suitable geometric realization they are in fact represented by simplicial prisms (simplest cases: parallelogram ($p = 1$); a triangular prism ($p = 2$)). The boundary relation is, when $\sigma_1 = ab$,

$$F(ab \times C_p) = b \times C_p - a \times C_p - ab \times F(C_p). \tag{1.1}$$

In a geometric realization the segments $ab \times$ (points of C_p), define a deformation (translation) of $a \times C_p$ into $b \times C_p$; the chains $ab \times$ (cells of C_p), are then called *deformation-cells*. In the abstract case they are the *deformation-elements* and $ab \times C_p$ is a *deformation-chain*. Writing generically C_p, C'_p for $a \times C_p$ (initial position), $b \times C_p$ (final position) and $\mathfrak{D}C_p$ for $ab \times C_p$, (1.1) is replaced by

$$F(\mathfrak{D}C_p) = C'_p - C_p - \mathfrak{D}F(C_p), \tag{1.2}$$

which is formula 3 of *Topology*, Chapter II.

2. Let $K = \{E^i_p\}$ be an abstract complex, and $L = \{C^i_p\}$ a collection of chains of K such that

$$F(C^i_p) = \eta^i_j C^j_{p-1}. \tag{2.1}$$

Let us introduce between the elements C the incidence relations $<$ defined by $C^i_p < C^j_p$, $C^j_{p-1} < C^i_p$ whenever $\eta^i_j \neq 0$, and all the incidences which result from them by imposing transitivity on the relation $<$. The chains incident with C^i_p shall be called the *faces* of C^i_p. Under the circumstances L is an abstract complex called a *chain-complex* in K. There may well exist identical relations between the elements of L, say:

$$x^i_j C^j_p = 0. \tag{2.2}$$

The group of the p-chains of L, to be generally designated by $G_p(L)$, is the additive group with the relations (2.2).

As an example if C_p is any chain of K, the chains C_p, $F(C_p)$ form a collection L and therefore they are the elements of a certain chain-complex K.

In conformity with general usage, the dimension of a chain-complex L is the maximum dimension of the chains in L.

3. Let $K = \{E_p^i\}$, $K' = \{E_p'^i\}$ be two complexes (chain-complexes are not excluded). By a transformation $T:K \to K'$ we understand an assignment of an $E' = TE$ to each E such that: (a) dim $TE \leqq$ dim E; (b) T transforms incident elements into incident elements. The transformation T induces a chain-transformation T_0 which is a collection of homomorphisms $G_p(K) \to G_p(K')$ determined by: $T_0E = TE$ when dim TE = dim E, $T_0E = 0$ otherwise. Conventionally we agree that $T_0C_{-1} = C_{-1}$.

Let us suppose now that K, K' are subcomplexes of the same abstract complex K^*, and let there be associated with T a collection of homomorphisms $G_p(K) \to G_{p+1}(K^*)$ such that for every C_p of K we have

$$F(\mathfrak{D}C_p) = T_0C_p - C_p - \mathfrak{D}F(C_p), \tag{3.1}$$

or symbolically

$$F\mathfrak{D} + \mathfrak{D}F = T_0 - 1. \tag{3.2}$$

We call the couple $\theta = (T, \mathfrak{D})$, a *chain-deformation* of K into K' over K^* (see DJ1). T is the transformation $K \to K'$, T_0 the chain-transformation induced by θ. We say of K, or of any one of its chains, that it is *chain-deformed* onto K' over K^*.

The aggregate of chains $K + K' + \sum \mathfrak{D}E$ is a chain-complex $\mathfrak{D}\mathfrak{K}$, the *deformation-complex* of K. From (3.1) follows that if Γ_p is a cycle of K then

$$F(\mathfrak{D}\Gamma_p) = T_0\Gamma_p - \Gamma_p,$$

so that

$$T_0\Gamma_p \sim \Gamma_p \text{ on } K^*.$$

In other words, *the extreme positions of a cycle which is chain-deformed over K^* are homologous on K^*.*

4. Theorem 1. *The chain-transformation T_0 induced by the chain-deformation $\theta: K \to K'$ is permutable with F. Conversely if $K = \{E_p^i\}$, $K' = \{E_p'^i\}$ are two abstract complexes and T is a transformation $K \to K'$ which induces a chain-deformation T_0 permutable with F, there exists an abstract complex $K^* \supset K + K'$ and a deformation $\theta: K \to K'$ over K^* which induces T.*

In the converse we are assuming that K, K' are abstract complexes, since we have to create K^* by adjunction of new elements which are not in any preassigned complex.

Let θ be the chain-deformation already considered. If we apply F to both sides of (1.2) we find

$$0 = FF\mathfrak{D}C = F\theta C - FC - F\mathfrak{D}FC. \tag{4.1}$$

From (1.2) applied to FC we deduce also

$$F\mathfrak{D}FC = \theta FC - FC, \tag{4.2}$$

and hence $\theta F = F\theta$ which proves the direct assertion.

The proof of the converse is more laborious. Let the equations of T_0 be

$$T_0 E_p^i = x_{p,j}^i E_p'^j. \tag{4.3}$$

We shall determine K^* so as to have just enough elements for our purpose. At all events it must contain certain chains $\mathfrak{D}C_p$ for which (1.2) holds. We set therefore

$$\mathfrak{D}(y_i E_p^i) = y_i \mathfrak{D}E_p^i, \tag{4.4}$$

and impose

$$F(\mathfrak{D}E_p^i) = T_0 E_p^i - E_p^i - \mathfrak{D}(F(E_p^i)), \qquad p > 0; \tag{4.5}$$

$$F(\mathfrak{D}E_0^i) = T_0 E_0^i - E_0^i. \tag{4.6}$$

If every $\mathfrak{D}E_q$, $q < p$, is known, (4.5) yields the boundaries of the p-elements. If the right-hand side in (4.5) is the boundary of some $(p+1)$-chain in $K + K'$, we take this chain, or one of them if there are several, as $\mathfrak{D}E_p^i$. In the contrary case we introduce a new abstract element $\mathfrak{D}E_p^i$ for which we postulate (4.5). At the same time we define as the faces of this new element besides itself: E_p^i, the elements $E_p'^j$ such that $x_{p,j}^i \neq 0$, together with all the faces of these elements. K^* consists of K, K' and all the new elements thus introduced. Of the three conditions for abstract complexes (Tucker loc. cit., also Bull. MS) the only one which is not fulfilled by construction is $FF = 0$, so that we must verify it. This needs to be done only for the new $\mathfrak{D}$ elements. First of all $FF(E_0^i) = 0$, since $TF = FT$ yields for the Kronecker index: $(T_0 E_0^i) = T_0(E_0^i) = (E_0^i)$. Proceeding by induction let us assume that we have verified $FF(\mathfrak{D}C_q) = 0$ for every $q < p$. The same for $q = p$ requires that

$$F(T_0 E_p^i) - F(E_p^i) - F(\mathfrak{D}F(E_p^i)) = 0. \tag{4.7}$$

From (4.5), together with the hypothesis of the induction we deduce:

$$F(\mathfrak{D}F(E_p^i)) = T_0 F(E_p^i) - F(E_p^i), \tag{4.8}$$

and, owing to $TF = FT$, this reduces to (4.7). This proves our theorem.

5. Application. Let K be an abstract simplicial complex and let a_i be its vertices ranged in a certain order. Consider $ab \times K$ and identify K with $a \times K$, denote $b \times K$ by K' and the vertices $b \times a_i$ by a_i'. Write every $\sigma_p = a_{i_0} \cdots a_{i_p}$ of K with the i's in ascending order and consider the chains

$$\mathfrak{D}\sigma_p = \sum (-1)^q a_{i_0} \cdots a_{i_q} a_{i_q}' \cdots a_{i_p}'. \tag{5.1}$$

By imposing homomorphism this defines $\mathfrak{D}$ for every chain of K, and we verify identically that (3.1) holds. Thus the isomorphism $T:K \to K'$ determined by $a_i \to a_i'$, together with $\mathfrak{D}$, determine a chain-deformation $K \to K'$ which we describe as a *translation*. This is justified on the ground that when the elements are now all made geometric so that $ab \times \sigma$, $ab \times K$, $\cdots$ are all prisms, the simplexes in the sum (5.1) form a simplicial decomposition of the prism $ab \times \sigma_p$.[3]

A noteworthy property is that in the present instance the simplexes making up $\mathfrak{D}\sigma$ have their vertices among those of σ and of $T\sigma$. We shall denote here by $\mathfrak{D}K$ the sum of the simplexes in (5.1) together with all their faces.

Suppose now that we have a simplicial complex K^* with subcomplexes K^1, K^2, and let T be a transformation $K^1 \to K^2$. Under our definitions T is simplicial and hence the induced chain-transformation T_0 commutes with F. There exists therefore an associated abstract chain-deformation θ in some complex $\supset K^1 + K^2$, but not necessarily in K^*. In order that the chain-deformation θ be possible in K^* a sufficient condition is that if $\sigma_p = a^1_{i_0}, \cdots, a^1_{i_p} \epsilon K^1$, (the indices are in increasing order) and if $Ta^1_i = a^2_i$, then all the simplexes $a^1_{i_0} \cdots a^1_{i_q} a^2_{i_q} \cdots a^2_{i_p}$ belong to K^*. For then we have manifestly in K^* a simplicial image of $\mathfrak{D}K$ above, with K isomorphic with K^1, and K' mapped into K^2. The images of the chains (5.1) will then be chains, say $\mathfrak{D}\sigma$, of K^*, which satisfy the basic boundary relations (3.2) demanded by θ. If the complexes were all geometric θ would determine a deformation $K^1 \to K^2$. We may describe it here as an abstract *deformation*, or more simply a deformation $K^1 \to K^2$ over K^*. In the particular case where T is one-one, θ is an *isotopy*.

II. $\mathfrak{P}$-Chains

6. We shall develop a theory of chains for a compact, metric space $\mathfrak{R}$. It will be convenient to assume for the present that $\mathfrak{R}$ has no isolated points. This restriction shall be removed later. We shall assume that we have a basic additive abelian group $\mathfrak{G}$, which is to play the rôle of coefficient-group.

Let $\mathfrak{K}$ be the complex consisting of all abstract simplexes σ whose vertices are points of $\mathfrak{R}$. The diameter of any σ is by definition the diameter of its set of vertices. By an ϵ complex of $\mathfrak{R}$, an ϵ chain of $\mathfrak{R}$ over $\mathfrak{G}$, we understand with Vietoris, a finite subcomplex of $\mathfrak{K}$, or a finite chain of $\mathfrak{K}$ over $\mathfrak{G}$ whose mesh $< \epsilon$.

Let K, L be two Vietoris complexes and let L be a transform of K under τ (necessarily simplicial). The Vietoris simplexes described in No. 5 required to have an (abstract) deformation $K \to L$, i.e. the simplexes of the chains $\mathfrak{D}\sigma$, $\sigma \epsilon K$, are all present. Therefore τ is induced by a deformation θ. The vertices of the simplexes of $\mathfrak{D}\sigma$ are all among those of σ and $\tau\sigma$. Therefore if

[3] The proof is given in a forthcoming paper in Fundamenta Mathematicae. See also regarding these questions *Topology* pp. 76–80, where deformations were first tied up explicitly with translations.

l.u.b. diam $(\sigma + \tau\sigma)$ for all σ, is $< \epsilon$, that is, if τ is an ϵ-transformation, τ will be an ϵ deformation also. In short, as regards the passage from K to L, ϵ transformations and deformations are wholly equivalent.

7. Consider now an infinite sequence of Vietoris complexes $\{K^i\}$, where K^i is finite, of mesh $< \epsilon_i$ and in addition:

(a) $\epsilon_i \to 0$ with $1/i$;

(b) there is a deformation $\pi_i: K^{i+1} \to K^i$;

(c) if we set $\pi_i^j = \pi_{j-1} \cdots \pi_i$, $i < j$, then for every $\epsilon > 0$ there is an n such that for $i, j > n$, π_i^j is an ϵ deformation $K^j \to K^i$.

It follows from (c) that for sufficiently large i, $\epsilon_i < \epsilon$, and also that $\pi_i (= \pi_i^{i+1})$ is then itself an ϵ deformation. Thus for i above a certain value the deformations π_i are all ϵ deformations. The π's are also called *projections* and $\{K^i\}$ a projection-sequence.

Let now $c_q = \{C_q^i\}$ be a sequence such that C_q^i is a chain of K^i over $\mathfrak{G}$, and $\pi_i C_q^{i+1} = C_q^i$. We call c_q a $\mathfrak{P}$-*chain of* $\mathfrak{R}$ *over* $\mathfrak{G}$; the chains C_q^i are the *components* of c_q and q is the *dimension* of the chain. A special q-dimensional chain is one in which all the C's are zero (have all coefficients zero), in which case we write $c_q = 0$. For $q = -1$ according to our conventions (No. 3), $\pi_i C_{-1}^{i+1} = C_{-1}^i = g$, so that a c_{-1} consists of an element $g \in \mathfrak{G}$ indefinitely repeated. All the $\mathfrak{P}$-chains thus obtained corresponding to a given g are considered as constituting one and the same c_{-1}, and denoted by g.

It is an elementary matter to verify that $\{-C_q^i\}$ and $\{F(C_q^i)\}$ are likewise $\mathfrak{P}$-chains with the same projection-sequence as c_q. They are respectively called the *negative* and the *boundary* of c_q, denoted by $-c_q$ and $F(c_q)$, and their dimension is q and $q - 1$. Moreover $FF(c_q) = 0$, since every $FF(C_q^i) = 0$.

To the chain c_q we attach the closed set $|c_q|$, consisting of all the points x such that every open set $U \supset x$ contains vertices of the chains C_q^i, where i is as high as we please. Explicitly also $|c_{-1}| = 0$. We have: $|-c_q| = |c_q|$, $|F(c_q)| \subset |c_q|$.

Kronecker-index:

When $q = 0$, $c_0 = \{C_0^i\}$, we may introduce a Kronecker-index (c_0) in the following way. Since $\pi_i C_0^{i+1} = C_0^i$, we have for the Kronecker-indices (sum of the coefficients of the vertices) the relation: $(C_0^{i+1}) = (C_0^i)$. Therefore (C_0^i) is independent of i, and its value is by definition (c_0). Under our conventions $F(C_0^i) = (C_0^i) = (c_0)$, and hence $F(c_0)$ is the (-1) dimensional chain c_{-1} generated by the element (c_0) of $\mathfrak{G}$ indefinitely repeated, or $F(c_0) = (c_0)$. In other words the same boundary relation subsists here as for simplicial chains.

Chain-complexes:

Consider the aggregate $\mathfrak{K} = \{c_q^i\}$ of all the $\mathfrak{P}$-chains on $\mathfrak{R}$. If we define: (a) the incidences between them by: $c_q^i < c_q^i$, $F(c_q^i) < c_q^i$, plus all the relations resulting therefrom by making $<$ transitive; (b) the incidence numbers [:] (Tucker's notation) by $[c:F(c)] = 1$, and all the others zero, $\mathfrak{K}$ becomes an

abstract complex. By a chain-complex L of $\mathfrak{R}$ we shall mean a subcomplex of $\mathfrak{R}$. Meshes, chain-deformations, ϵ or otherwise, said to be on $\mathfrak{R}$, are then defined in the natural way. The complex L is thus a chain-complex in the sense of No. 2, except that the (-1)-chains and Kronecker-indices may now be elements of $\mathfrak{G}$ and not merely integers. The q-chains of L, $q \geqq 0$, are still assumed with integer coefficients. The group $\mathfrak{G}$ enters in them indirectly through the components C_q^i of the chains.

8. Chain-groups. Our first difficulties arise with the definition of *addition*. We shall find it convenient in fact to distinguish two kinds of additions, the one of geometric type, the other more algebraic.

Suppose that we have two chains $C_q = \{C_q^i\}$, $c_q' = \{C_q'^i\}$, associated with the same projection-sequence $\{K^i\}$. Then $\{C_q^i + C_q'^i\}$ is another chain with the same projection-sequence. We call it *geometric sum* of c_q, c_q' and designate it by $c_q + c_q'$ in the usual manner. It is also immediately apparent that

$$-(c_q + c_q') = (-c_q) + (-c_q'), \tag{8.1}$$

$$F(c_q + c_q') = F(c_q) + F(c_q'). \tag{8.2}$$

The definition of addition just given is the most natural. It suffers however from the grave defect that the sum $c_q + c_q'$ is only defined for certain special pairs c_q, c_q' and hence it may not be used to form an additive group comprising all the q-chains as elements. We must therefore look for a somewhat different addition. It will be in fact an addition of chain-classes which are based on a certain notion of equivalence to be developed presently.

9. Let $\{K^i\}$, $\{K'^i\}$ be two projection-sequences with projections π_i, π_i' and let there exist, for each i, a deformation $\theta_i : K^i \to K'^i$ whose amount $\eta_i \to 0$ with $1/i$, and where in addition $\theta_i \pi_i = \pi_i' \theta_{i+1}$. We call the aggregate $\theta = \{\theta_i\}$ a *regular deformation* of the first projection-sequence into the second, a *regular isotopy* whenever every θ_i is isotopic. In the latter case the inverses θ_i^{-1} are defined and isotopic and $\{\theta_i^{-1}\}$ defines a regular isotopy, the *inverse* of θ and denoted by θ^{-1}. Similarly if $\theta' = \{\theta'^i\}$ is a second regular deformation $\{K'^i\} \to \{K''^i\}$, $\{\theta'^i \theta^i\}$ defines a regular deformation $\{K^i\} \to \{K''^i\}$ the product of the two and denoted by $\theta'\theta$. Finally if every $\theta_i = 1$, we write $\theta = 1$.

Let now c_q be the same chain as previously, attached to $\{K^i\}$, and let θ be a regular deformation $\{K^i\} \to \{K'^i\}$. If $C_q'^i = \theta_i C_q^i$, we have $C_q'^i = \theta_i \pi_i C_q^{i+1} = \pi_i' C_q'^{i+1}$, so that $c_q' = \{C_q'^i\}$ is a $\mathfrak{B}$-chain attached to $\{K'^i\}$. We shall denote this new chain by θc_q. We notice that under our conventions $\theta_i g = g$, so that $\theta c_{-1} = c_{-1}$. Moreover since $\theta_i F = F\theta_i$, $\theta_i(-C^i) = -\theta_i C^i$, we have $\theta F = F\theta$, $\theta(-c) = -\theta c$, and for $q = 0$, $\theta(c_0) = (c_0)$. Finally since θ_i is an ϵ_i deformation, where $\epsilon_i \to 0$ with $1/i$, we have $|\theta c_q| = |c_q|$, or θ does not modify the sets associated with the $\mathfrak{B}$-chains.

Let now $c_q = c_q^1, c_q^2, \cdots, c_q^r = c_q'$, be a finite sequence wherein one of any two consecutive terms, (we do not say which) is derived from the other by a regular deformation θ. We then say that c_q is *equivalent* to c_q' and write $c_q \simeq c_q'$. This equivalence is symmetric and transitive, likewise reflexive, since we may choose $\theta = 1$. Therefore it leads to a partition of the chains into classes to be denoted by $[c_q]$. The set $|c_q|$ depends solely on the class $[c_q]$. We observe also that if c_q remains in a fixed class so do $-c_q$ and $F(c_q)$. The two classes thus determined are designated by $-[c_q]$ and $F[c_q]$. Finally, since $(\theta c_0) = (c_0)$, the chains of a class $[c_0]$ all have the same Kronecker-index, to be denoted by $([c_0])$.

It is a simple matter to show that if $c_q' = \{C_q'^i\}$ is a chain with a projection-sequence $\{K'^i\}$ such that the first is a subsequence of $\{C_q^i\}$ and the second the corresponding subsequence of $\{K^i\}$ with projections π_i' which are the same projections as between elements of the latter, then $c_q \simeq c_q'$.

10. Let now $c_q, c_q', \cdots$ be as before, save that the chains need not be equivalent. Let x_j, x_j' be the vertices of K^i, K'^i. Since $\mathfrak{R}$ has no isolated points, we may find in the respective $\frac{1}{2^i}$-neighborhoods of x_j, x_j' points $\bar{x}_j, \bar{x}_j'$ which are all distinct. Let θ, θ' denote the isotopies $x \to \bar{x}$, $x' \to \bar{x}'$, and let $\theta_i K^i = \bar{K}^i, \cdots$. It is clear that $\bar{c}_q = \{\bar{C}_q^i\}$, $\bar{c}_q' = \{\bar{C}_q'^i\}$ are $\mathfrak{P}$-chains and that $c_q \simeq \bar{c}_q'$, $c_q' \simeq \bar{c}_q'$. Now however $\{\bar{K}^i + \bar{K}'^i\}$ is a projection-sequence with $\bar{C}_q^i, \bar{C}_q'^i$ as chains of its ith complex. Therefore $\bar{c}_q + \bar{c}_q'$ exists, is well defined and equal to $\bar{c}_q' + \bar{c}_q$. If $\theta^*, \theta^{*\prime}$ are other isotopies analogous to θ, θ', leading to $K^{*i}, \cdots$ then τ equal to $\theta^*\theta^{-1}$ on $\{\bar{K}^i\}$ and to $\theta^{*\prime}\theta'^{-1}$ on $\{\bar{K}'^i\}$, is a regular isotopy $\{\bar{K}^i + \bar{K}'^i\} \to \{K^{*\prime} + K^{*\prime i}\}$, from which follows that $c_q^* + c_q^{*\prime} \simeq \bar{c}_q + \bar{c}_q'$. Therefore the class $[\bar{c}_q + \bar{c}_q']$ is fixed. Practically the same argument shows that if $c_q \simeq d_q$, then $\bar{c}_q, \bar{d}_q, \bar{c}_q'$ may be obtained as before so that $\bar{c}_q + \bar{c}_q'$ and $\bar{d}_q + \bar{c}_q'$ exist and are equivalent. Similarly with c, c' interchanged since $+$ is commutative. Replacing thus first c, then c' by equivalent chains we find that $[\bar{c}_q + \bar{c}_q']$ depends solely upon $[c_q]$ and $[c_q']$. This new class is called the *sum* of the other two and denoted by $[c_q] \dot{+} [c_q']$.

Let c_q, c_q' be both associated with the same projection-sequence $\{K^i\}$ and let $\bar{c}_q, \bar{c}_q'$ be as before. There are then regular isotopies $\theta^{-1}: \bar{c}_q \to c_q$, $\theta'^{-1}: \bar{c}_q' \to c_q'$. Since $\bar{K}^i, \bar{K}'^i$ have no common vertex there is a regular deformation $\tau: \{\bar{K}^i + \bar{K}'^i\} \to \{K^i\}$ such that $\tau_i = \theta_i^{-1}$ on $\bar{K}^i$, $= \theta_i'^{-1}$ on $\bar{K}'^i$ and $\tau(\bar{c}_q + \bar{c}_q') = c_q + c_q'$. Therefore whenever the chains have a geometric sum, the class of the latter is the sum of their classes.

Finally we have by construction

$$F([c_q] \dot{+} [c_q']) = F[c_q] \dot{+} F[c_q']; \qquad -([c_q] \dot{+} [c_q']) = (-[c_q]) \dot{+} (-[c_q']).$$

Let γ_q denote a $\mathfrak{P}$-chain whose components C_q^i are all zero. We verify immediately that $[\gamma_q] = [-\gamma_q]$ and $[c_q] \dot{+} [\gamma_q] = [c_q]$. Therefore in particular all

chains γ_q form a unique class such that if we denote it by 0, it has all the properties of the zero relative to the addition $\dotplus$ of classes of q-chains.

From previous results we have also that whatever c_q, $FF[c_q] \simeq 0$.

11. If we combine our results on the classes of q-chains we find that they form relative to the law of combination $\dotplus$, and with the zero just defined, an additive abelian group Γ_q such that F is a homomorphism $\Gamma_q \to \Gamma_{q-1}$ for every q with $FF = 0$. Γ_q is not yet our final chain-group, for in its present form it does not enable us to take account of the dimension of $\mathfrak{R}$. However we are very near the goal.

Let us define as a *null-class*, and denote henceforth by 0, any class $[c_q]$ containing an element c_q such that there exists for every i an η_i deformation τ_i of C_q^i into a chain $\tau_i C_q^i$ which is degenerate or zero, where $\eta_i \to 0$ with $1/i$. When this occurs we shall now write $[c_q] \simeq 0$. If $[c_q]$ is null so is $-[c_q]$ (obvious) and also $F[c_q]$ (since the boundary of a degenerate simplicial chain is degenerate or zero). In particular also $[FF(c_q)]$ is a null-class whatever c_q. When $q = 0$ if $[c_0] \simeq 0$ we have $([c_0]) = 0$; since its value is $F[c_0]$ and the latter is null only when equal to zero. Regarding the sum, if $[c_q] \simeq 0$, $[c_q'] \simeq 0$ so is $[c_q] \dotplus [c_q']$. For let c_q, c_q' be the chains in their respective classes with η_i deformations τ_i, τ_i' of C_q^i, $C_q'^i$, as described above, into degenerate or zero-chains, where $\eta_i \to 0$ with $1/i$. Let also θ, θ' be the regular isotopies of No. 10 of the chains to chains $\bar{c}_q$, $\bar{c}_q'$ such that $\bar{c}_q + \bar{c}_q'$ is well defined. If we define $\bar{\tau}_i$ as $\tau_i \theta_i^{-1}$ on $\bar{C}_q^i$ and $\tau_i' \theta_i'^{-1}$ on $\bar{C}_q'^i$, τ_i is a ζ_i deformation of $\bar{C}_q^i + \bar{C}_q'^i$ into a degenerate or zero-chain, where $\zeta_i \to 0$ with $1/i$. Therefore $[\bar{c}_q + \bar{c}_q^i] \simeq 0 \simeq [c_q] \dotplus [c_q']$.

We conclude then that the null-classes form a subgroup Δ_q of Γ_q such that $F\Delta_q \subset \Delta_{q-1}$. Therefore the difference group $G_q = \Gamma_q - \Delta_q$ has the property that FG_q is a subgroup of G_{q-1}. *These groups G_q are the groups which we call the groups of $\mathfrak{P}$-chains of $\mathfrak{R}$ over $\mathfrak{G}$, and we shall denote them by $G_q(\mathfrak{R})$.*

Conclusion:

If we identify all the chains of an equivalence class, and all the classes equivalent modulo a null-class, the operation $\dotplus$ on the chains leads to the usual type of additive chain-groups and therefore also to the usual homology groups for the $\mathfrak{P}$-chains of $\mathfrak{R}$ over $\mathfrak{G}$.[4]

---

[4] Whenever $\mathfrak{G}$ is topological the homology groups H_q may be so topologized as to become a topological space. The procedure merely requires that we define suitable neighborhoods of the zero in H_q and the following will do: a neighborhood U of the zero consists of all the homology classes containing a cycle $\gamma_q = \{\Gamma_q^i\}$ such that for i above a certain n: (a) the coefficients of Γ_q^i are in a specified neighborhood V of the zero in $\mathfrak{G}$; (b) mesh $\Gamma_q^i < \epsilon$, where ϵ is a specified positive number. Whenever the topologization is possible we may define as homology groups in place of the groups H_q their closures $\bar{H}_q$. There are important questions regarding the universal group of coefficients for $\mathfrak{P}$-cycles, such as those considered by N. E. Steenrod, American Journal of Mathematics, vol. 58 (1936), pp. 661–701. We shall not discuss them here.

An example:

It may be shown by an example that $\mathfrak{P}$-chains are essentially different from singular chains. $\mathfrak{R}$ consists of the following planar set: (a) the segment $x = 0$, $-1 \leqq y \leqq 1$; (b) $0 < x \leqq 1$, $y = \sin 1/x$; (c) a broken line joining $(0, 1)$ to $(1, 0)$ and not meeting the two preceding sets anywhere else. If the $\mathfrak{P}$-chains are taken with rational coefficients the set has a Betti-number $R_1 = 1$, whereas in the system of rational singular chains $R_1 = 0$. Indeed there are no singular one-cycles over any group of coefficients $\mathfrak{G}$ whatsoever.

12. We may now remove the restriction as to the isolated points of $\mathfrak{R}$ imposed in No. 6. In any case if there are isolated points, each is both an open and a closed set. Therefore the sum $\mathfrak{R}_i$ of the isolated points is open and its complement, the dense part $\mathfrak{R}_d$ of $\mathfrak{R}$, is closed. Moreover since $\mathfrak{R}$ is separable $\mathfrak{R}_i$ is countable. Let $y_1, y_2, \cdots$ be its points. We now define the typical $c_q = \{C_q^h\}$ by the condition that $C_q^h = C_q'^h + C_q''^h$, where $c_q' = \{C_q'^h\}$ is a chain of $\mathfrak{R}_d$ and $c_q'' = \{C_q''^h\}$ is a chain of a finite subset α of $\mathfrak{R}_i$, independent of h. The geometric sum may be defined as before, and clearly $c_q = c_q' + c_q''$. When $q > 0$ we define equivalence and class relations as those determined by the components c_q' alone, that is to say we consider c_q'' as null throughout. When $q = 0$, we have for h above a certain values $C_0''^h = g^j y_j$, $g^j \epsilon \mathfrak{G}$, the sum being finite. We then define two chains $c_0 = c_0' + c_0''$, $d_0 = d_0' + d_0''$ as equivalent when $c_0' \simeq d_0'$ in the previous sense on $\mathfrak{R}_d$ and $c_0'' = d_0''$. We define $-c_0$ as the chain composed of $-c_0''$ and $-d_0'$ and if $c_0'' = g^j y_j$, $d_0'' = \bar{g}^j y_j$, we define $[c_0] \dotplus [d_0]$ as the class determined by a chain of $[c_0'] \dotplus [d_0']$ (relative to $\mathfrak{R}_d$), together with $(g^j + \bar{g}^j) y_j$, and the rest is more or less obvious.

In practice we shall not need the class notation [ ]. In all questions of equivalence a chain c_q is to stand merely for a representative of its whole class, and $c_q \simeq 0$ shall mean that c_q is in a null-class.

13. THEOREM 2. *If* $\dim \mathfrak{R} = n$, *all* q*-chains with* $q > n$, *are equivalent to zero.*

Let first $\mathfrak{F} = \{F^i\}$ be a finite covering of $\mathfrak{R}$ by closed sets and let λ, μ be its Lebesgue constant and mesh, where $\lambda \leqq 2\mu$. Choose for each F^i a point A^i on the set, and let Φ = nerve $\mathfrak{F}$ based on the vertices A^i. Let K be any finite Vietoris-complex whose mesh $< \lambda$ and let τ be the transformation $K \to \Phi$ determined by the condition that if B is any vertex of K, then τB is a point A^i such that $F^i \supset B$. As is well known τ is simplicial and if σ is any element of K then diam $(\sigma + \tau\sigma) <$ mesh $\Phi + 2\mu < 4\mu$. Therefore all complexes whose mesh $< \lambda$ may be 4μ deformed onto Φ.

Consider now a sequence of coverings $\mathfrak{F}^i$ such as $\mathfrak{F}$, where μ_i (= mesh $\mathfrak{F}^i$) $\to 0$ with $1/i$, $\mathfrak{F}^1 = \mathfrak{R}$, and order $\mathfrak{F}^i \leqq n = \dim \mathfrak{R}$. Thus the nerves Φ^i constructed as above will all be n dimensional at most and Φ^1 = a point. We also have λ_i = Lebesgue constant of $\mathfrak{F}^i \leqq 2\mu_i$, and hence $\lambda_i \to 0$ with $1/i$.

Let now $c_q = \{C_q^i\}$ with $\{K^i\}$ as its projection-sequence. We define a trans-

formation τ_h for each K^h as follows: (a) For all K^h whose mesh $\geqq \lambda_2$, we choose $\tau_h: K^h \to \Phi^1$; (b) for all K^h whose mesh $\geqq \lambda_{i+1}$ and $< \lambda_i$, there is a simplicial $4\mu_i$ transformation $K^h \to \Phi^i$, and we choose it as τ_h. Since the transformed complexes $\tau_h K^h$ are of dimension $\leqq n$, when $q > n$ the chains $\tau_h C_q^h$ are all degenerate or zero. Therefore $c_q \simeq 0$.

14. Subdivision. Let $\mathfrak{R} = \{c_q^i\}$, $\mathfrak{L} = \{d_q^i\}$ be two chain-complexes of $\mathfrak{R}$. We say that $\mathfrak{L}$ is a subdivision of $\mathfrak{R}$ whenever:

(a) every chain of $\mathfrak{R}$ is a geometric sum of chains of $\mathfrak{L}$:

$$c_q^i = \xi_{q,j}^i d_q^j, \qquad \xi_{q,j}^i = \pm 1, 0; \tag{14.1}$$

(b) in (14.1) whenever $\xi_{q,j}^i \neq 0$ we have $| d_q^j | \subset | c_q^i |$;

(c) no proper subcomplex of $\mathfrak{L}$ possesses the two preceding properties.

It is hardly necessary to point out that the ξ's can only be $\neq 0$ for a collection of chains admitting geometric addition.

Let $\mathfrak{R}$ be finite and suppose that we have a finite chain-complex $\mathfrak{L}$ for which only (a), (b) hold. If (c) does not hold for $\mathfrak{L}$ then it has a proper subcomplex $\mathfrak{L}^1$ for which (a) and (b) still hold, etc. We thus have a sequence $\mathfrak{L}, \mathfrak{L}^1, \cdots$, each term of which is a proper subcomplex of its predecessor, and such that (a) and (b) hold for all terms. Since the number of elements in $\mathfrak{L}^i$ is finite and decreases monotonely with increasing i, the sequence is finite. Its last term, $\mathfrak{L}'$, satisfies (a), (b), and (c) and hence it is a subdivision of $\mathfrak{R}$. Therefore to prove that a finite $\mathfrak{R}$ may be subdivided we merely need to obtain a finite complex $\mathfrak{L}$ satisfying (a), (b).

In accordance with the customary terminology an ϵ subdivision is one whose mesh $< \epsilon$. We shall now construct for every finite $\mathfrak{R}$ an ϵ subdivision with noteworthy properties to be described as we go.

Let $\mathfrak{R}$ be finite and let dim $\mathfrak{R} = p$. Choose any p-chain c_p^i of $\mathfrak{R}$ and let $\{K^{ih}\}$ be its projection-sequence with π_i as the projections. Let $\{\sigma^m\}$ be the simplexes of K^{ih} ranged in some order. If ζ is any simplex of C_p^{ij}, $j \geqq h$, the successive simplexes $\zeta, \pi_{j-1}^j \zeta, \cdots, \pi_h^j \zeta$ are simplexes of $K^{ij}, K^{i,j-1}, \cdots, K^{ih}$ with the same projection σ^m in K^{ih}. It follows that we may write

$$C_p^{ij} = \sum C_p^{ijm} \tag{14.2}$$

where $c_p^{im} = \{C_p^{ijm}\}$, $j = 1, 2, \cdots$, is a $\mathfrak{P}$-chain such that $\pi_h^j C_p^{ijm}$ is a p-chain of σ^m. We shall have therefore (geometric summation):

$$c_p^i = \sum c_p^{im}. \tag{14.3}$$

Clearly by construction $| c_p^{im} | \subset | c_p^i |$. Let us operate similarly with every c_p^i and the same h, and let all the chains c_p^{im} thus obtained be designated in some order by d_p^t.

The aggregate $\{d_p^t, F(d_p^t), c_q^i\}$, $q < p$, fulfills conditions (a), (b) relative to p alone and also conditions (a), (b) of No. 2, so that by inserting the proper

incidences and incidence-numbers we may turn it into a chain-complex $\mathfrak{L}^1$ fulfilling (a), (b) as to p. We now treat $\mathfrak{L}^1$ in the same way relative to the same integer h and its $(p-1)$-chains and obtain thus $\mathfrak{L}^2 = \{d_p^s, d_{p-1}^t, F(d_{p-1}^t), c_q^i\}$, $q < p-1$, behaving like $\mathfrak{L}^1$, and in particular satisfying (a), (b) for the dimensions p, $p-1$. The repetition of this process $p+1$ times leads to a chain-complex $\mathfrak{L} = \{d_q^s\}$ which is finite and fulfills (a), (b). It possesses therefore a subcomplex $\mathfrak{L}'$ which is a subdivision of $\mathfrak{K}$. For the sake of simplicity we shall assume that $\mathfrak{L}$ is already the subdivision.

If $d_q^s = \{D_q^{sj}\}$, it follows from our construction that D_q^{sh} is a chain of one of the simplexes of $K^h = \sum K^{ih}$. Let $\{\sigma^m\}$ denote this time these simplexes ranged in some order. The transformation $\rho: d_q^s \to D_q^s$ is a simplicial chain-transformation which commutes with F. This transformation ρ preserves the geometric identities between the chains. That is to say if

$$g_s d_q^s = 0 \tag{14.4}$$

then also

$$g_s \rho d_q^s = 0. \tag{14.5}$$

For (14.4) implies that all the chains properly appearing in the relation are associated with the same projection-sequence $\{K^j\}$ and that

$$g_s D_q^{sj} = 0 \tag{14.6}$$

for every j. Taking $j = h$ this becomes precisely (14.5).

Since K is finite h may be chosen so large that all the projections π_h^j involved in the initial projection-sequences $\{K^{ij}\}$ are all ϵ projections. We shall then have for $j > h$: diam $(D_q^{sj} + \rho d_q^s) < \epsilon$. Hence from the definition of the sets $|c|$: diam $(|d_q^s| + \rho d_q^s) < \epsilon$. Therefore ρ is then an ϵ transformation of the subdivision $\mathfrak{L}$ into the Vietoris complex K^h.

If we combine all our results we have then:

THEOREM 3. *Every finite chain-complex $\mathfrak{K}$ of $\mathfrak{R}$ has, for every ϵ, a subdivision $\mathfrak{L}$ which may undergo an ϵ transformation permutable with F into a Vietoris complex K. Both mesh K and mesh $\mathfrak{L}$ are $< \epsilon$.*

Any chain c_p is a chain of a chain complex $\mathfrak{K}$, for example the one which may be derived from the pair c_p, $F(c_p)$ after the manner of No. 2. If we apply our theorem to this complex we have the

COROLLARY. *Given any chain c_p and any $\epsilon > 0$, we may find chains c_p^i such that*: (a) $c_p = \xi_i c_p^i$; (b) $|c_p^i| \subset |c_p|$; (c) diam $|c_p^i| < \epsilon$; (d) *the aggregate $\{c_p^i\}$ may undergo an ϵ transformation permutable with F into a Vietoris chain.*

III. APPLICATION TO LOCAL CONNECTEDNESS

15. As the situation now stands it is not essentially difficult to extend to local connectedness based on $\mathfrak{P}$-chains the known properties of what we have termed HLC spaces (DJ1). We shall modify our notations to make them

somewhat more elastic. Instead of LC, $\overline{\text{LC}}$ we shall write LC*, LC**. The first are the absolute neighborhood retracts, the second the absolute retracts. LC will be reserved for the whole class of spaces with local connectedness in the sense of homotopy as $n-$LC, LC^n, $\cdots$, of Ann. M. 1, and they shall be called *geometric* LC. The similar classes where in place of homotopy we have chain-deformations, or if we prefer, chains and cycles of a class κ in place of singular cells and spheres, shall be denoted by $n-\text{LC}(\kappa)$ $\cdots$, and called *algebraic* LC. In particular we have $n-\text{LC}(\mathfrak{P})$, $\cdots$. We shall indicate the salient points in the theory of these last spaces, reserving complete proofs for another occasion.

Let $\mathfrak{K}$, $\mathfrak{L}$ be two finite closed abstract complexes with $\mathfrak{L}$ a subcomplex of $\mathfrak{K}$. We shall say that $\mathfrak{L}$ is *dense* in $\mathfrak{K}$ whenever every basis element of $\mathfrak{K}$ has a face in $\mathfrak{L}$. In particular this implies that every basis element, which is a cycle, is in $\mathfrak{L}$. When $\mathfrak{K}$ is simplicial $\mathfrak{L}$ is dense when and only when it includes all the vertices of $\mathfrak{K}$.

Let $L = \{c_p^i\}$ be a transform, under T, of $\mathfrak{L}$ in $\mathfrak{R}$, where the c's are $\mathfrak{P}$-chains. We call L a *partial-realization* of $\mathfrak{K}$ in $\mathfrak{R}$. Any element C of $\mathfrak{K}$ is represented by one or more faces, say c'^i, in L. The mesh of L as a partial realization is l.u.b. diam $\sum |c'^i|$ for all $C \in \mathfrak{K}$. Let it be possible to extend T from $\mathfrak{L}$ to $\mathfrak{K}$, to a transformation $\overline{T}$ ($\overline{T} = T$ on $\mathfrak{L}$), and let K denote the chain-complex on $\mathfrak{R}$ which is the image of $\mathfrak{K}$ under $\overline{T}$. Evidently L is a subcomplex of K. We say then that the partial realization L of $\mathfrak{K}$ has been extended to a full realization K.

DEFINITION. *The space $\mathfrak{R}$ is said to be* LC*($\mathfrak{P}$) *whenever* for every $\epsilon > 0$ there is an $\eta(\epsilon)$ such that every partial realization whose mesh $< \eta(\epsilon)$ may be extended to a full realization whose mesh $< \epsilon$. The space is said to be LC**($\mathfrak{P}$) whenever in addition every partial realization may be extended to a full realization (without any restrictions regarding the meshes).

If the extension condition holds only for complexes $\mathfrak{K}$ whose dimension $\leqq n+1$, $\mathfrak{R}$ is said to be LC^n($\mathfrak{P}$). It may be shown as in Ann. M. 1 that the LC^n($\mathfrak{P}$) type is characterized as follows: for every ϵ there is an η such that every q-cycle, $q \leqq n$, whose diameter $< \eta$ bounds a chain whose diameter $< \epsilon$.

16. The basic theorem in this theory is:

THEOREM 4. *When $\mathfrak{R}$ is* LC^n($\mathfrak{P}$) *every chain-complex K on $\mathfrak{R}$ whose dimension $\leqq n$, has a subdivision which is ϵ chain-deformable onto a fixed finite simplicial chain-complex Ψ whose dimension $\leqq n$. When the space is* LC*($\mathfrak{P}$) *this holds regardless of dimensions. When $\mathfrak{R}$ is* LC**($\mathfrak{P}$) *every K is chain-deformable onto a point.*

We shall merely outline the proof. Let $\mathfrak{F}$, A^i, Φ, λ, μ be as in No. 13 and let Φ_n be the maximal subcomplex of Φ of dimension $\leqq n$. The vertices A^i may be considered as a partial realization of Φ_n. For μ small enough we may obtain the full realization Ψ. There is a subdivision K' of K which may undergo

a ν transformation permutable with F into a simplicial complex and if $\nu < \lambda$, the complex may be transformed into a subcomplex K'' of Ψ. The ultimate transformation $K \to K''$ being permutable with F, $K + K''$ is a partial realization of a certain product $\sigma_1 \times \mathfrak{R}$, where if $\sigma_1 = ab$, $a \times \mathfrak{R}$ has for image K and the mesh of the partial realization may be made as near mesh Φ + mesh Ψ as we please. Therefore for μ small enough the full extension may be carried out and its existence proves the theorem. For an LC*($\mathfrak{P}$) we merely have $\Phi_n = \Phi$, but the rest is the same. Finally when $\mathfrak{R}$ is LC**(P) the mesh restrictions may be disregarded and hence we may take $\Phi = \Psi =$ a point.

From this theorem follows readily the

COROLLARY. *The homology groups of an* LC^n($\mathfrak{P}$) *for dimentions* $\leqq n$ *are subgroups of those of a definite finite complex of dimension* $\leqq n$. *Similarly for all the groups of an* LC*($\mathfrak{P}$), *without reference to dimension, except insofar that all the groups above a certain dimension vanish. The groups of an* LC**($\mathfrak{P}$) *are those of a point.*

We may also prove by means of the subdivision theorem the

THEOREM 5. *When* $\mathfrak{R}$ *is* LC^n($\mathfrak{P}$) *its homology groups for the dimensions* $\leqq n$ *based on* $\mathfrak{P}$*-cycles are isomorphic with the corresponding Vietoris groups. When* $\mathfrak{R}$ *is* LC*($\mathfrak{P}$) *this holds for all dimensions. Therefore the groups of the Corollary may be taken to be the Vietoris groups.*

By means of Theorem 3 we may also prove:

THEOREM 6. *When* dim $\mathfrak{R} \leqq n$, *and* $\mathfrak{R}$ *is* LC^n($\mathfrak{P}$) *it is* LC*($\mathfrak{P}$).

Fixed points

We have recently given a very simple proof of our fixed point formula for geometric LC. Let us merely state that it applies with insignificant modifications to LC*($\mathfrak{P}$), and indeed to any algebraic LC* space based on chains which are adequately subdivisible. For example it holds when the algebraic LC* are based on singular chains.

PRINCETON UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ON PRODUCTS IN A COMPLEX[1]

BY HASSLER WHITNEY

(Received June 10, 1937; Revised November 9, 1937)

1. Introduction

In classical homology theory, founded by Poincaré, the fundamental operation is that of forming the boundary ∂A^p of a chain A^p. This is found by multiplying the coefficients of A^p into a matrix of incidence. Algebraically, an equally obvious operation, using the same matrix of incidence, forms the "coboundary" δA^{p-1} from a given A^{p-1}. It has recently been discovered that the algebraic part of the theory of intersections of chains in a manifold, when interpreted with the other operation, could be generalized to arbitrary complexes. It is the object of this paper to give a complete treatment of the fundamentals of this theory. We use a general type of complex and general coefficient groups, and prove the required invariance theorems. Parts of the paper are new only in form. Various notions used here appear first in Tucker's thesis, [12].[2]

In Part I we define, after Tucker, the complexes to be used. The cohomology groups are defined, and then elementary properties of "dual homomorphisms" are given. The latter are used throughout the paper.

Suppose we ask for a product of p-chains A^p and q-chains B^q, giving $(p+q)$-chains $A^p \smile B^q$, which shall have topological significance. A p-cell times a q-cell far away from the p-cell should certainly give nothing; hence (P_1) of §5 is a natural assumption. Considering δ as the fundamental operation, if we wish the multiplication to give a result in the cohomology groups, we must have cocycle $\smile$ cocycle = cocycle. Hence $\delta(A \smile B)$ must be expressible in terms such as $\delta A \smile B$ and $A \smile \delta B$. (P_2) is the natural form.[3] Suppose we ask that a vertex times itself equal itself. (Hence the γ of §5 is 1.) Then (using Theorem 1), at least in an ordinary connected complex, the products exist, and when carried out in the cohomology groups, are uniquely determined (Theorem 5). This may be considered the fundamental theorem of the present paper.

---

[1] Presented to the American Mathematical Society, under a different title, March 27, 1937. An outline of the paper appeared in Proc. Nat. Ac. Sci., vol. 23 (1937), pp. 285–291.

[2] The numbers in square brackets refer to the bibliography at the end of the paper.

[3] Note that the dimension in each term is correct. It is easily seen that the signs must alternate in one of the terms. We might call $\smile$ and $\frown$ "cup" and "cap". Equivalent formulas (see also (Q_2) and (19.2)) occur in a fundamental manner in the classical theory; see for instance Lefschetz, [10], pp. 111, 169 and 226. The same formula appears in the theory of differential forms.

Another product, $A^p \frown B^q = C^{q-p}$, is considered; it is algebraically equivalent to the $\smile$ product. We also consider briefly a particular definition of the products which may be used in simplicial complexes. The rest of Part II is devoted to proving other elementary properties and to showing how general coefficient groups may be used.

In Part III we answer two questions. First, in a polyhedron, are homology and cohomology groups and products independent of the particular simplicial subdivision chosen? The proof §16 that they are (which may be considered well-known) is relatively simple; the considerations show that the groups and products may be associated with an abstract space. Secondly, to find the groups and products, it is often very inconvenient to have to use simplicial subdivisions; but we must then show that a general complex gives the same theory as a simplicial subdivision. This combinatorial theorem (Theorem 14) occupies the rest of Part III. It turns out that complexes which may be used for determining the groups often may not be used for determining the products.[4] However, most of the invariance proof may be carried out in the general type of complex. Similar (but slightly weaker) theorems for homology groups have been proved by Tucker, [12], and Alexandroff-Hopf, [4], Ch. VI.

The relation of the two products to intersection theory in a manifold is considered briefly in Part IV. In contrast with Čech, [5], we use the classical method of dual complexes. In Part V, the products are considered in product complexes and in Euclidean space. As an application of preceding results, some mapping theorems are proved, due in part to H. Hopf.

Some special topics are considered in the Appendix.

Historical note. The coboundary of a chain, when a passage to the limit is applied, becomes the derived of a covariant alternating tensor (compare Alexander, [1]). In this form, of course, it has long been known. From the algebraic standpoint, cocycles appear in a different form in the "pseudocycles" of S. Lefschetz, [10]. Cocycles may be interpreted as cycles in the "dual complex," considered in papers by W. Mayer[5] and A. W. Tucker, [12]. An application of cocycles in their direct form was given in our note on Sphere-spaces.[6]

The work of L. Pontrjagin on character groups led to the realization that not only the homology but also the cohomology groups might be important. At the International Topological Conference, Moscow, 1935, J. W. Alexander and A. Kolmogoroff presented papers giving not only the theory of cohomology groups (with different notations), but also defining a product (for simplicial

---

[4] For instance, in a torus, for determining the groups, we may subdivide into one vertex, two 1-cells, and one 2-cell; but for determining the products, we must have say four vertices, eight 1-cells, and four 2-cells. (Or we may use Part V.)

[5] Monatsh. f. Math. u. Phys., vol. 36 (1929), pp. 1–42 and 219–258.

[6] Proc. Nat. Ac. of Sci., vol. 21 (1935), pp. 464–468, §6.

complexes) in the groups.[7] It appears that D. van Danzig[8] and E. Čech also had a portion of these results. However, the Kolmogoroff-Alexander product was not wholly satisfactory, it being too large by a numerical factor. In studying their product at the end of 1935, the author discovered the $\smile$ product of §6. In an effort to generalize a theorem of H. Hopf (see footnote 17), the remaining results of §6 and §7 were found.[9] At the same time, Čech discovered the same products; see [5]. Alexander studied the revised $\smile$ product; see [3]. Finally, the present Part II may be considered an outgrowth of Čech's paper [5]. Our assumptions (P_1) and (P_2) are closely allied to those of Čech, §2. Our proof of Theorem 5 was obtained after a study of a corresponding proof in Čech's paper. In the mean time, H. Freudenthal, [6], also found the results of §7, and studied the relation of these products to other known products.

Recently S. Lefschetz[10] has shown that the products of §6, when properly translated into the residual space of a sphere containing the complex, become the classical intersections. As in Tucker, [12], he gives postulates which a "chain-product" should satisfy. Neither author proves a uniqueness theorem (such as our Theorem 5).

I. Preliminaries

2. The complexes used

Complexes, in topology, are certain algebraic structures which may be given a geometric significance. For a given algebraic structure to be geometrically realizable, certain conditions must be satisfied. For instance, we may demand that there be a simplicial subdivision (algebraically defined); any simplicial complex determines a geometric complex in Euclidean space. Complexes of this nature we shall say "admit a simplicial subdivision" (see Part III). To define a product theory in the complex, stronger conditions are necessary; a complex satisfying these conditions "admits a product theory." Most of the paper will be concerned with these complexes.

A *complex K admitting a product theory* is a system as follows. It has cells[11]

---

[7] See [2], [8] and [9]. The Kolmogoroff-Alexander product is not that given in [2].

[8] See Recueil Math., Moscow, vol. 1 (43), (1936), pp. 672–674. Cohomology groups with general coefficient groups are studied by Steenrod, Am. Journal of Math., vol. 58 (1936), pp. 661–701.

[9] These results were applied in classifying the maps of a 3-sphere into a 2-sphere; see Bull. Am. Math. Soc., vol. 42 (1936), p. 338. They were communicated in letters to L. Zippin early in 1936.

[10] Bull. Am. Math. Soc., vol. 43 (1937), pp. 345–359, §5, (d). It should be noted that without a postulate such as our (P_1), the product may give pratically any product in the cohomology groups.

[11] We assume the cells are finite in number; however, most of the results extend to the infinite case, at least with the proper definitions.

σ_i^p of dimension p. There are two types of relations. A cell may be a *face* of a cell of higher dimension, and is a face of itself. Two cells σ_i^{p-1} and σ_j^p of neighboring dimension have an *incidence number* ${}^p\partial_j^i$, which is an integer. In terms of these numbers we define boundaries ∂A^p etc. as usual (see §3). The *closure* $\bar{\sigma}_i^p$ of a cell σ_i^p is the set of all its faces. The *star* St (σ) of σ is the set of cells with σ as a face. The closure and star of any subcomplex are defined similarly. We say a cycle $A^p = \sum \alpha_i \sigma_i^p$ is *boundary-like* if $p > 0$ or if $p = 0$ and $\sum \alpha_i = 0$ (or $I \cdot A^1 = 0$; see (2.1) and (3.2)). The definition as here given is useful only in the case that δI contains no 1-cells which have two vertices as faces; see Lemma 5.

We make the following assumptions.

(K_1) If σ_1 is a face of σ_2 and σ_2 is a face of σ_3, then σ_1 is a face of σ_3.

(K_2) If ${}^p\partial_j^i \neq 0$, then σ_i^{p-1} is a face of σ_j^p.

(K_3) $\partial\partial A^p = 0$ always (or equivalently, $\delta\delta A^p = 0$).[12]

(K_4) Each boundary-like cycle (with integer coefficients) in any $\bar{\sigma}_i^p$ bounds a chain in $\bar{\sigma}_i^p$.

Note that, using (4.7) below, (K_3) is equivalent to the vanishing (for all A^p, B^{p-2}) of any of $\partial\partial A^p \cdot B^{p-2}$, $\partial A^p \cdot \delta B^{p-2}$, $A^p \cdot \delta\delta B^{p-2}$, or $\delta\delta B^{p-2}$.

Certain elementary properties of these complexes are the following:

LEMMA 1. *If $p > 0$, then $\partial\sigma_i^p \neq 0$.*

For if $\partial\sigma_i^p = 0$, then σ_i^p is a cycle in $\bar{\sigma}_i^p$. By (K_4), it must bound a $(p+1)$-chain in $\bar{\sigma}_i^p$. But $\bar{\sigma}_i^p$ contains no $(p+1)$-cells.

LEMMA 2. *Every cell of dimension > 0 has a lower dimensional face, and hence a vertex as a face.*

This follows from the last lemma and (K_2) and (K_1).

LEMMA 3. *Each 1-cell has just one or two vertices as faces.*

By the last lemma, we need merely prove that any σ^1 has not three vertices a, b, c as faces. If it had, then by (K_4), $b - a = k\partial\sigma^1$ and $c - b = l\partial\sigma^1$. Hence $l(b - a) = k(c - b)$, which is impossible (as $k \neq 0$).

The definitions of open and closed subcomplexes are as usual, in terms of "being a face of"; for σ_1 and σ_2 in the subcomplex K' of K, we say σ_2 is a face of σ_1 in K' if it is in K.

LEMMA 4. *If K admits a product theory, so does every closed subcomplex.*

The proof is simple.

Note that, if

$$I = \sum \sigma_i^0 = \text{sum of all the vertices of } K, \tag{2.1}$$

we may have $\delta I \neq 0$. But see Theorem 2. If K is connected, and $\delta I = 0$, the only 0-cocycles are the multiples of I.

---

[12] Complexes satisfying (K_1), (K_2) and (K_3) are exactly those considered by Tucker, [12]. If (K_4) is satisfied also, and $\delta I = 0$ (that is, K is "augmentable"), then Tucker shows that K admits a simplicial subdivision and hence is geometrically realizable.

LEMMA 5. *If each* 1*-cell is on two vertices, then* $\delta I = 0$ *and* K *admits a simplicial subdivision* (see §13).

For take any 1-cell σ; say $\partial\sigma = \alpha a + \beta b$. As $a - b$ is boundary-like, there is a k such that $a - b = k\partial\sigma$; hence $\beta = -\alpha$. Therefore $\delta I \cdot \sigma = I \cdot \partial\sigma = \alpha + (-\alpha) = 0$, and $\delta I = 0$. The last statement has been proved by Tucker (see footnote 12).

3. Homology and cohomology groups[13]

The *boundary* and *coboundary* of chains are defined by

$$(3.1)\quad \partial\left(\sum_i \alpha_i \sigma_i^p\right) = \sum_{i,j} \alpha_i{}^p \partial_i^j \sigma_j^{p-1}, \qquad \delta\left(\sum_i \alpha_i \sigma_i^p\right) = \sum_{i,j} \alpha_i{}^{p+1} \partial_j^i \sigma_j^{p+1}.$$

The α_i are integers or elements of an abelian group G.[14] The chain A^p is a *cycle* [*cocycle*] if $\partial A^p = 0$ [$\delta A^p = 0$]. We say

A is *homologous to* B, $A \sim B$, if $A - B$ is a boundary,
A is *cohomologous to* B, $A \backsim B$, if $A - B$ is a coboundary.

As $\partial\partial A^p = 0$, $\delta\delta A^p = 0$ (see (K_3)), we may define as usual the homology and cohomology groups. Using the coefficient group G, we denote these by ${}^p\mathbf{H}^G$, ${}^p\mathbf{H}_G$. The group of p-chains of a complex with integer coefficients will be denoted by L^p.

We define the *scalar product* of two chains of the same dimension by

$$(3.2)\quad \left(\sum \alpha_i \sigma_i^p\right) \cdot \left(\sum \beta_i \sigma_i^p\right) = \sum \alpha_i \beta_i.$$

Note that $A^p \cdot \sigma^p$ is the coefficient of σ^p in the chain A^p.

4. Dual homomorphisms[15]

Let G and G' be free groups with *fixed* sets of generators $a_1, \cdots, a_p$ and $a_1', \cdots, a_q'$. Then to any matrix of integers $\| \phi_{ij} \|$ $(i = 1, \cdots, p, j = 1, \cdots, q)$ correspond homomorphisms of G into G' and of G' into G, defined by

$$(4.1)\quad \phi a_i = \sum_j \phi_{ij} a_j', \qquad \phi' a_i' = \sum_j \phi_{ji} a_j.$$

If ϕ is any homomorphism of G into G', then $\phi a_i = \sum \phi_{ij} a_j'$ for some integers ϕ_{ij}; thus the matrix $\| \phi_{ij} \|$ is defined, and hence the homomorphism ϕ'. Each of ϕ, ϕ' determines the other uniquely; the matrix of one is the transposed matrix of the other. We call these homomorphisms *dual*, and write

$$(4.2)\quad \phi' = D(\phi), \qquad \phi = D(\phi') = D(D(\phi)).$$

---

[13] For further details, see for instance Whitney, [13].

[14] Except where otherwise stated, we use integers: $G = I_0$.

[15] Compare the Appendix, §27. Dual homomorphisms correspond to adjoint linear transformations in algebra. They have been used in topology by Tucker, [12], §25.

If ϕ maps G into G' and ψ maps G' into G'', then $\theta = \psi\phi$ maps G into G''. Its matrix is clearly

$$\| \theta \| = \| \psi \| \, \| \phi \| : \qquad \theta_{ij} = \textstyle\sum_k \psi_{ik}\phi_{kj} .$$

On transposing, we find

$$\theta' = D(\theta) = D(\psi\phi) = D(\phi)D(\psi) = \phi'\psi'. \tag{4.3}$$

Linear combinations of homomorphisms are defined as usual. We find

$$D(c\phi + d\psi) = cD(\phi) + dD(\psi). \tag{4.4}$$

We shall apply these homomorphisms to the groups L^p, the individual cells forming the generators. Clearly *∂ and δ are dual.* If ϕ maps $L^p(K)$ into $L^{p+k}(K')$ for each p (k fixed), and $\phi' = D(\phi)$, then

$$\delta\phi' A' = \phi' \delta A' \text{ (all } A') \text{ if and only if } \partial\phi A = \phi\partial A \text{ (all } A). \tag{4.5}$$

For one relation follows from the other on taking duals. These relations hold in particular for simplicial maps. Using (3.2), we have

$$A \cdot \phi' B' = \phi A \cdot B' \text{ if and only if } \phi' = D(\phi). \tag{4.6}$$

For

$$\phi\sigma_i \cdot \sigma'_j = \left(\textstyle\sum_k \phi_{ik}\sigma'_k\right) \cdot \sigma'_j = \phi_{ij} , \qquad \sigma_i \cdot \phi'\sigma'_j = \phi'_{ji} .$$

Hence

$$A^p \cdot \delta B^{p-1} = \partial A^p \cdot B^{p-1}. \tag{4.7}$$

REMARK. The above definitions and results hold equally well if G and G' are vector spaces with fixed bases.

Finally, note that

$$A \cdot B = 0 \text{ if } \delta A = 0 \text{ and } B \sim 0 \text{ or if } A \backsim 0 \text{ and } \partial B = 0. \tag{4.8}$$

For if $\delta A = 0$ and $B = \partial C$, then $A \cdot B = A \cdot \partial C = \delta A \cdot C = 0$, etc. Hence, if $\delta A = 0$ and $B \sim B'$, then $A \cdot B = A \cdot B'$, etc.

II. THE PRODUCTS

5. Definition and properties of products

We shall use only integer coefficients until §11. Corresponding to each p-cell σ_i^p, q-cell σ_j^q, and $(p+q)$-cell σ_k^{p+q}, we wish to find an integer ${}^{pq}\Gamma_k^{ij}$, such that the following properties hold.

(Γ_1). If σ_i^p and σ_j^q are not both faces of σ_k^{p+q}, then ${}^{pq}\Gamma_k^{ij} = 0$.

(Γ_2). For all p, q, i, j, k,

$$\textstyle\sum_l {}^{p+q+1}\partial_k^l \, {}^{pq}\Gamma_l^{ij} = \sum_l {}^{p+1}\partial_l^i \, {}^{p+1,q}\Gamma_k^{lj} + (-1)^p \sum_l {}^{q+1}\partial_l^j \, {}^{p,q+1}\Gamma_k^{il}.$$

(Γ_3). For some integer γ, and all q and j,

$$\sum_i {}^{0q}\Gamma_j^{ij} = \gamma.$$

The important case is the case $\gamma = 1$; after Part II, we always take $\gamma = 1$.

In terms of the quantities ${}^{pq}\Gamma_k^{ij}$, we define two bilinear operations on chains, as follows.

$$\sigma_i^p \smile \sigma_j^q = \sum_k {}^{pq}\Gamma_k^{ij}\sigma_k^{p+q}, \tag{5.1}$$

$$\sigma_j^q \frown \sigma_k^{p+q} = \sum_l {}^{pq}\Gamma_k^{lj}\sigma_l^p. \tag{5.2}$$

Clearly

$$(\sigma_i^p \smile \sigma_j^q)\cdot\sigma_k^{p+q} = {}^{pq}\Gamma_k^{ij} = \sigma_i^p\cdot(\sigma_j^q \frown \sigma_k^{p+q}), \tag{5.3}$$

and hence

$$(A^p \smile B^q)\cdot C^{p+q} = A^p\cdot(B^q \frown C^{p+q}). \tag{5.4}$$

Two products $\smile$ and $\frown$ correspond if and only if (5.4) holds. Given one of $\smile$, $\frown$, we can find the corresponding Γ by (5.3), and hence the other.

The above properties translate into the following, for the $\smile$ product.

(P_1) $\sigma_i^p \smile \sigma_j^q$ is a $(p+q)$-chain in $\mathrm{St}\,(\sigma_i^p)\cdot\mathrm{St}\,(\sigma_j^q)$.[16]

(P_2) $\delta(A^p \smile B^q) = \delta A^q \smile B^q + (-1)^p A^p \smile \delta B^q$.

(P_3) For some integer γ, $I \smile \sigma_j^q = \gamma\sigma_j^q$ (all σ_j^q).

Also, for the $\frown$ product,

(Q_1) $\sigma_j^q \frown \sigma_k^{p+q}$ is a p-chain in $\overline{\mathrm{St}\,(\sigma_j^q)\cdot\sigma_k^{p+q}}$.

(Q_2) $\partial(A^q \frown B^{p+q}) = (-1)^p\delta A^q \frown B^{p+q} + A^q \frown \partial B^{p+q}$.

(Q_3) For some integer γ, $I\cdot(\sigma_j^q \frown \sigma_j^q) = \gamma$ (all σ_j^q).

We shall prove the equivalence of the three sets of properties. (P_1) is clearly equivalent to (Γ_1). If we write (P_2) with σ_i^p and σ_j^q, use (5.1), and consider the coefficient of σ_k^{p+q+1} on each side, we obtain (Γ_2); conversely, (Γ_2) implies (P_2). (Γ_3) and (P_3) are clearly equivalent. Suppose (Γ_1) holds; we shall prove (Q_1). If σ_j^q is not a face of σ_k^{p+q}, then ${}^{pq}\Gamma_k^{ij} = 0$ for every i, hence $\sigma_j^q \frown \sigma_k^{p+q} = 0$, and ($Q_1$) holds. If σ_j^q is a face of σ_k^{p+q}, then $\mathrm{St}\,(\sigma_j^q)\cdot\sigma_k^{p+q} = \sigma_k^{p+q}$; by ($\Gamma_1$), $\sigma_j^q \frown \sigma_k^{p+q}$ is a chain in $\bar{\sigma}_k^{p+q} = \overline{\mathrm{St}\,(\sigma_j^q)\cdot\sigma_k^{p+q}}$. Suppose ($Q_1$) holds. If σ_j^q is not a face of σ_k^{p+q}, then $\mathrm{St}\,(\sigma_j^q)\cdot\sigma_k^{p+q} = 0$, $\sigma_j^q \frown \sigma_k^{p+q} = 0$, and ${}^{pq}\Gamma_k^{ij} = 0$ (all i). If σ_i^p is not a face of σ_k^{p+q}, then (Q_1) shows that it does not occur in $\sigma_j^q \frown \sigma_k^{p+q}$; hence ${}^{pq}\Gamma_k^{ij} = 0$. (Q_2) is seen to be equivalent to (Γ_2), if we replace p by $p+1$. Finally, (Q_3) is equivalent to (Γ_3).

[16] $S\cdot T$ is the subcomplex of K containing those cells in both S and T. Thus $\mathrm{St}(\sigma_1)\cdot\sigma_2 = \sigma_2$ if σ_1 is a face of σ_2, and $= 0$ otherwise. Čech, [5], assumes merely that the chain is in $\mathrm{St}(\sigma_i^p)$.

THEOREM 1. *If K is connected and $\delta I = 0$, then (Γ_3) is a consequence of (Γ_1) and (Γ_2).*

By Lemmas 1 and 2, it is sufficient to show (using the (P_i)) that if

$$\partial\sigma^q = \alpha\sigma^{q-1}\ (\alpha \neq 0), \qquad I \smile \sigma^{q-1} = \gamma\sigma^{q-1}, \qquad I \smile \sigma^q = \gamma'\sigma^q,$$

then $\gamma = \gamma'$. (That $I \smile \sigma^h = \theta\sigma^h$ for some θ follows from (P_1).) By (P_2),

$$\delta(I \smile \sigma^{q-1}) = I \smile \delta\sigma^{q-1} = I \smile \alpha\sigma^q + \cdots = \alpha\gamma'\sigma^q + \cdots;$$

also

$$\delta(I \smile \sigma^{q-1}) = \delta\gamma\sigma^{q-1} = \gamma\alpha\sigma^q + \cdots,$$

as $\alpha \neq 0, \gamma = \gamma'$.

THEOREM 2. For all σ_j^q, $\delta I \smile \sigma_j^q = 0$.

For, by (P_2) and (P_3),

$$\delta I \smile \sigma_j^q = \delta(I \smile \sigma_j^q) - I \smile \delta\sigma_j^q = \delta\gamma\sigma_j^q - \gamma\delta\sigma_j^q = 0.$$

THEOREM 3. *If $\delta I = 0$, then*

$$I \smile \sigma^p = \sigma^p \smile I = I \frown \sigma^p = \gamma\sigma^p \qquad \text{(all } \sigma^p\text{)}. \tag{5.5}$$

First, $\sigma^0 \smile I = \sigma^0 \smile \sigma^0 = I \smile \sigma^0 = \gamma\sigma^0$. Suppose $\sigma^{p-1} \smile I = \gamma\sigma^{p-1}$ and $\delta\sigma^{p-1} = \alpha\sigma^p + \cdots$ $(\alpha \neq 0)$. Clearly $\sigma^p \smile I = \theta\sigma^p$ for some θ. Then as $\delta I = 0$, (P_2) and (P_3) give

$$\delta(\sigma^{p-1} \smile I) = \delta\sigma^{p-1} \smile I = \alpha\theta\sigma^p + \cdots,$$

$$\delta(\sigma^{p-1} \smile I) = \delta(\gamma\sigma^{p-1}) = \gamma\alpha\sigma^p + \cdots,$$

and $\theta = \gamma$. That $\sigma^p \smile I = \gamma\sigma^p$ now follows, by Lemma 1. The last relation is proved similarly, considering $\partial(I \frown \sigma^p)$.

THEOREM 4. *The $\smile$ and $\frown$ products define products among the cohomology and homology groups, thus*:

cohomology class $\smile$ cohomology class = cohomology class,

cohomology class $\frown$ homology class = homology class.

Explicit formulas are given in (11.6) and (11.7). That cocycle $\smile$ cocycle = cocycle, cocycle $\frown$ cycle = cycle, follows from (P_2) and (Q_2). Also

$$A_1 \smile B \backsim A_2 \smile B \quad \text{if} \quad A_1 \backsim A_2 \quad \text{and} \quad \delta B = 0, \tag{5.6}$$

$$A \smile B_1 \backsim A \smile B_2 \quad \text{if} \quad B_1 \backsim B_2 \quad \text{and} \quad \delta A = 0, \tag{5.7}$$

$$A_1 \frown B \sim A_2 \frown B \quad \text{if} \quad A_1 \backsim A_2 \quad \text{and} \quad \partial B = 0, \tag{5.8}$$

$$A \frown B_1 \sim A \frown B_2 \quad \text{if} \quad B_1 \sim B_2 \quad \text{and} \quad \delta A = 0. \tag{5.9}$$

For instance, if $A_2 - A_1 = \delta C$ and $\partial B = 0$, then by (Q_2),

$$A_2 \frown B - A_1 \frown B = \delta C \frown B = \pm\partial(C \frown B).$$

Hence the definitions in the theorem are unique.

THEOREM 5. *For each integer γ, the products of Theorem 4 exist in any complex and are uniquely determined.*

This will be proved in §8 and §9. The relation between products with different γ's is given in Theorems 9 and 10.

In §10 we shall prove the two *associative* laws and the *commutative* law:

If K admits a simplicial subdivision (see §13), *then*

$$A \smile (B \smile C) \backsim (A \smile B) \smile C \quad \text{if} \quad \delta A = \delta B = \delta C = 0, \tag{5.10}$$

$$A \frown (B \frown C) \sim (A \smile B) \frown C \quad \text{if} \quad \delta A = \delta B = \partial C = 0, \tag{5.11}$$

it being understood in (5.11) that $\frown$ and $\smile$ have the same γ. *If $\delta I = 0$, then*

$$A^p \smile B^q \backsim (-1)^{pq} B^q \smile A^p \quad \text{if} \quad \delta A^p = \delta B^q = 0. \tag{5.12}$$

THEOREM 6. *Let K and K' be simplicial, let f be a simplicial map of K into K', and let f' be the dual of f. Then*

$$f'(A' \smile B') \backsim f'A' \smile f'B' \quad \text{if} \quad \delta A' = \delta B' = 0, \tag{5.13}$$

$$f(f'A' \frown B) \sim A' \frown fB \quad \text{if} \quad \delta A' = \partial B = 0. \tag{5.14}$$

These generalize a theorem of Hopf to arbitrary complexes.[17] For the proof, see §10. For the use of different coefficient groups, see §11.

6. The products in simplicial complexes

If K is simplicial, a very simple definition of Γ, $\smile$, $\frown$ is possible. Order the vertices of K in a fixed manner. Each simplex σ^p may now be written in the normal form $\sigma^p = x_{i_0} \cdots x_{i_p}$, $i_0 < \cdots < i_p$. We define, if $x_{i_0} \cdots x_{i_p} \cdots x_{i_{p+q}}$ is a simplex,

$$^{pq}\Gamma^{i_0 \cdots i_p, i_p \cdots i_{p+q}}_{i_0 \cdots i_p \cdots i_{p+q}} = 1 \qquad (i_0 < \cdots < i_p < \cdots < i_{p+q}), \tag{6.1}$$

and $\Gamma = 0$ for any other triple of simplexes. (The meaning of the Γ should be clear.) In terms of $\smile$ and $\frown$, this gives for instance (if $x_1x_2x_3x_4$ is a simplex)

$$x_1x_2 \smile x_2x_3x_4 = x_1x_2x_3x_4, \qquad x_2x_3x_4 \frown x_1x_2x_3x_4 = x_1x_2.$$

(Γ_1) and (Γ_3) obviously hold. A simple calculation gives (Γ_2); see Alexander, [4].

For these products we clearly have

$$A \smile (B \smile C) = (A \smile B) \smile C, \quad A \frown (B \frown C) = (A \smile B) \frown C. \tag{6.2}$$

7. Products and simplicial maps

Having ordered the vertices of K', order those of K so that

$$\text{if } f(x_i) = x'_{\theta(i)}, \text{ then } \theta(i) < \theta(j) \text{ implies } i < j. \tag{7.1}$$

---

[17] H. Hopf, [7], Satz I.

Using the products of §6, we shall prove (5.13) and (5.14) in the form:

THEOREM 7. *In simplicial complexes, using the products of* §6,

$$(7.2) \qquad f'(A' \smile B') = f'A' \smile f'B', \qquad f(f'A' \frown B) = A' \frown fB.$$

To prove the first equation, take any $\sigma_i'^p$ and $\sigma_j'^q$; say

$$\sigma_i'^p = x_0' \cdots x_p', \qquad \sigma_j'^q = x_p' \cdots x_{p+q}';$$

if the vertices of the simplexes are not related as shown, then clearly both sides of the equation (with these simplexes) vanish. By definition of f',

$$f'\sigma_i'^p = \sum x_{\alpha_0} \cdots x_{\alpha_p}, \qquad \theta(\alpha_h) = h,$$
$$f'\sigma_j'^q = \sum x_{\beta_p} \cdots x_{\beta_{p+q}}, \qquad \theta(\beta_h) = h.$$

Hence

$$f'\sigma_i'^p \smile f'\sigma_j'^q = \sum x_{\alpha_0} \cdots x_{\alpha_p} \cdots x_{\alpha_{p+q}}, \qquad \theta(\alpha_h) = h.$$

This is clearly also $f'(\sigma_i'^p \smile \sigma_j'^q)$.[18] To prove the second equation, we use the first, (4.6) and (5.4): For any cell σ' (of the proper dimension),

$$\sigma' \cdot f(f'A' \frown B) = f'\sigma' \cdot (f'A' \frown B) = (f'\sigma' \smile f'A') \cdot B = f'(\sigma' \smile A') \cdot B$$
$$= (\sigma' \smile A') \cdot fB = \sigma' \cdot (A' \frown fB).$$

8. Construction of the products

In a general complex, it is most convenient to construct the $\frown$ product. Take any integer γ. We shall construct all $\sigma_i^q \frown \sigma_j^{p+q}$ in succession for $p = 0, 1, 2, \cdots$. First consider $p = 0$. Set $\sigma_i^q \frown \sigma_j^q = 0$ for $j \neq i$. Let $\sigma_j^q \frown \sigma_j^q$ be any 0-chain in $\bar{\sigma}_j^q$ the sum of whose coefficients is γ; this is possible, by Lemma 2. The required properties hold so far. Suppose all $\sigma_i^q \frown \sigma_j^{r+q}$ are properly constructed for all q and all $r < p$; we must construct each $\sigma_i^q \frown \sigma_j^{p+q}$. If σ_i^q is not a face of σ_j^{p+q}, set $\sigma_i^q \frown \sigma_j^{p+q} = 0$. As $\mathrm{St}(\sigma_i^q) \cdot \bar{\sigma}_j^{p+q}$ has no cells, both sides of (Q_2) with σ_i^q and σ_j^{p+q} vanish. Now suppose σ_i^q is a face of σ_j^{p+q}. Set

$$(8.1) \qquad C^{p-1} = (-1)^p \delta\sigma_i^q \frown \sigma_j^{p+q} + \sigma_i^q \frown \partial\sigma_j^{p+q};$$

this is a chain in $\bar{\sigma}_j^{p+q}$.

Suppose first that $p = 1$. Then for some α, $\delta\sigma_i^q = \alpha\sigma_j^{p+q} + \cdots$, and

$$(8.2) \qquad I \cdot C^{p-1} = -I \cdot (\alpha\sigma_j^{p+q} \frown \sigma_j^{p+q}) + I \cdot (\sigma_i^q \frown \alpha\sigma_i^q)$$
$$= -\alpha\gamma + \alpha\gamma = 0.$$

Hence, by (K_4), we may choose the 1-chain $\sigma_i^q \frown \sigma_j^{p+q}$ in $\bar{\sigma}_j^{p+q}$ so that its boundary is C^{p-1}. Then (Q_1) through (Q_3) hold.

---

[18] By using joins, and the 0-chains $X_h = f'x_h'$, one could give a direct formal proof.

Suppose next that $p > 1$. Then (Q_2) gives

$$\partial C^{p-1} = (-1)^p[(-1)^{p-1}\delta\delta\sigma_i^q \frown \sigma_j^{p+q} + \delta\sigma_i^q \frown \partial\sigma_j^{p+q}] + (-1)^{p-1}\delta\sigma_i^q \frown \partial\sigma_j^{p+q} + \sigma_i^q \frown \partial\partial\sigma_j^{p+q} = 0. \tag{8.3}$$

Hence, by (K_4), we may choose $\sigma_i^q \frown \sigma_j^{p+q}$ in $\bar{\sigma}_j^{p+q}$ with the boundary C^{p-1}. The properties again hold.

We remark that any $\frown$ product may be constructed in this manner.

9. Uniqueness of the products

We first prove

THEOREM 8. *Let $\frown$ be any product with $\gamma = 0$. Then there is a bilinear operation $\wedge$ such that*

(R_1) $\sigma_i^q \wedge \sigma_j^{p+q}$ is a $(p+1)$-chain in $\overline{\mathrm{St}(\sigma_i^q)\cdot\sigma_j^{p+q}}$.

(R_2) $\sigma_i^q \frown \sigma_j^q = \partial(\sigma_i^q \wedge \sigma_j^q)$.

(R_3) $\sigma_i^q \frown \sigma_j^{p+q} = \partial(\sigma_i^q \wedge \sigma_j^{p+q}) + (-1)^p\delta\sigma_i^q \wedge \sigma_j^{p+q} + \sigma_i^q \wedge \partial\sigma_j^{p+q}$ for $p > 0$.

We shall construct $\wedge$ for $p = 0, 1, \cdots$. As $\gamma = 0$, we can construct $\sigma_i^q \wedge \sigma_j^q$ so (R_2) holds. Suppose all $\sigma_i^q \wedge \sigma_j^{r+q}$ are constructed for all q and $r < p$; we shall construct $\sigma_i^q \wedge \sigma_j^{p+q}$. We make it 0 if σ_i^q is not a face of σ_j^{p+q}. Suppose it is. Set

$$C^p = \sigma_i^q \frown \sigma_j^{p+q} - (-1)^p\delta\sigma_i^q \wedge \sigma_j^{p+q} - \sigma_i^q \wedge \partial\sigma_j^{p+q}.$$

By (Q_2), and (R_3) solved for $\partial(\sigma_i^q \wedge \sigma_j^{p+q})$,

$$\partial C^p = (-1)^p\delta\sigma_i^q \frown \sigma_j^{p+q} + \sigma_i^q \frown \partial\sigma_j^{p+q} - (-1)^p[\delta\sigma_i^q \frown \sigma_j^{p+q} - \delta\sigma_i^q \wedge \partial\sigma_j^{p+q}] - [\sigma_i^q \frown \partial\sigma_j^{p+q} - (-1)^{p-1}\delta\sigma_i^q \wedge \partial\sigma_j^{p+q}] = 0.$$

Hence C^p is a cycle in $\bar{\sigma}_j^{p+q}$, and (as $p > 0$) we may choose $\sigma_i^q \wedge \sigma_j^{p+q}$ with C^p as boundary.

We may form linear combinations of $\frown$ products (and also of $\smile$ products) by defining

$$A(\alpha_1\frown_1 + \alpha_2\frown_2)B = \alpha_1(A \frown_1 B) + \alpha_2(A \frown_2 B). \tag{9.1}$$

By applying this to any $\sigma_j^q \frown \sigma_j^q$, we find

$$\gamma(\alpha_1\frown_1 + \alpha_2\frown_2) = \alpha_1\gamma(\frown_1) + \alpha_2\gamma(\frown_2). \tag{9.2}$$

THEOREM 9. *For any two products $\frown$ and $\frown'$,*

$$\gamma(\frown)(A^p \frown' B^q) \sim \gamma(\frown')(A^p \frown B^q) \quad \text{if} \quad \delta A^p = \partial B^q = 0. \tag{9.3}$$

As a consequence, if v^p and u^q denote cohomology and homology classes, then

$$\gamma(\frown)(v^p \frown' u^q) = \gamma(\frown')(v^p \frown u^q). \tag{9.4}$$

To prove this, set $\frown'' = \gamma(\frown)\frown' - \gamma(\frown')\frown$. Then $\gamma(\frown'') = 0$. By Theorem 8, (R_2) or (R_3), $A^p \frown'' B^q \sim 0$, from which (9.3) follows.

THEOREM 10. *Theorem* 9 *holds for the* $\smile$ *product. Thus*

$$\gamma(\smile)(A^p \smile' B^q) \backsim \gamma(\smile')(A^p \smile B^q) \quad \textit{if} \quad \delta A^p = \delta B^q = 0. \tag{9.5}$$

As above, it is sufficient to show that if $\gamma(\smile) = 0$, then $A^p \smile B^q \backsim 0$ for all cocycles A^p, B^q. Let $\frown$ correspond to $\smile$; then $\gamma(\frown) = 0$, and we may construct $\wedge$. Construct a corresponding $\vee$ (using relations of the form (5.1), (5.2)); then, as in (5.4),

$$(D^{p+1} \vee E^q)\cdot F^{p+q} = D^{p+1}\cdot(E^q \wedge F^{p+q}). \tag{9.6}$$

Now for $p > 0$ and any cell $\sigma = \sigma_k^{p+q}$, as $A\cdot\partial C = \delta A\cdot C$,

$$\begin{aligned}(A \smile B)\cdot\sigma = A\cdot(B \frown \sigma) &= A\cdot[\partial(B \wedge \sigma) + (-1)^p \delta B \wedge \sigma + B \wedge \partial\sigma] \\ &= (\delta A \vee B)\cdot\sigma + (-1)^p(A \vee \delta B)\cdot\sigma + \delta(A \vee B)\cdot\sigma,\end{aligned}$$

and hence

$$A^p \smile B^q = \delta(A^p \vee B^q) + \delta A^p \vee B^q + (-1)^p A^p \vee \delta B^q; \tag{9.7}$$

if $p = 0$, then $A^0 \smile B^q = \delta A^0 \vee B^q$. Hence if A and B are cocycles, then $A \smile B \backsim 0$.

10. Proof of properties in §5

To prove (5.12), we define a new product, $\smile'$, by

$$A^p \smile' B^q = (-1)^{pq} B^q \smile A^p.$$

Clearly (P_1) holds for $\smile'$. To prove (P_2), (P_2) for $\smile$ gives

$$\begin{aligned}\delta(A^p \smile' B^q) &= (-1)^{pq}[\delta B^q \smile A^p + (-1)^q B^q \smile \delta A^p] \\ &= (-1)^{pq+(q+1)p} A^p \smile' \delta B^q + (-1)^{pq+q+q(p+1)} \delta A^p \smile' B^q,\end{aligned}$$

which reduces to the desired formula. (P_3) is a consequence of Theorem 3. (5.12) now follows from (9.5).

To prove (5.10) and (5.11), take a simplicial subdivision K' of K, and define new products in K, using those of §6 in K', by (15.3) and (15.6). The properties now follow at once from (6.2). (Compare the proof of Theorem 14.)

To prove Theorem 6, we apply (9.5) and (7.2): Using $\smile_0$ for the product of §6,[19]

$$f'(A' \smile B') \backsim \gamma(\smile)f'(A' \smile_0 B') = \gamma(\smile)(f'A' \smile_0 f'B') \backsim f'A' \smile f'B'.$$

The other relation is proved similarly.

---

[19] If $C' \backsim D'$, then $D' - C' = \delta E'$, and

$$f'D' - f'C' = f'\delta E' = \delta f'E', \qquad f'C' \backsim f'D'.$$

11. The products, general coefficient groups

Let G be an abelian group. Let $\mathrm{Ch}_Y(X)$ be the group of homomorphisms (characters) of the group X into the group Y. Then $H = \mathrm{Ch}_G(G)$ is a ring with the definitions (writing $h \cdot g$ for $h(g)$)

$$(h + h') \cdot g = h \cdot g + h' \cdot g, \qquad (hh') \cdot g = h \cdot (h' \cdot g). \tag{11.1}$$

If we use the coefficient group G in chains with which we form cycles, then the most useful coefficients for chains from which cocycles are formed are those of H.[20]

EXAMPLES. Let I_0, I_m, R, $\mathfrak{R}$, R_1, $\mathfrak{R}_1$ denote the groups of integers, integers mod m, real numbers, rational numbers, reals mod 1, and rationals mod 1. All but the last two are rings. Then, using $\approx$ for isomorphic, examples of G and corresponding H are

(α) $G \approx I_0$, $H \approx I_0$.

(β) $G \approx I_m$, $H \approx I_m$ ($m \geqq 1$).

(γ) $G \approx R$ or $\mathfrak{R}$, $H \approx R$ or $\mathfrak{R}$.

(δ) $G \approx R_1$ or $\mathfrak{R}_1$, $H \approx I_0$.

(α) is a special case of (β). H is commutative in these cases. If G is a direct sum of such groups, we may find the corresponding H with the following rules:

$$\mathrm{Ch}_Y(X_1 + X_2) \approx \mathrm{Ch}_Y(X_1) + \mathrm{Ch}_Y(X_2).$$

$$\mathrm{Ch}_{Y_1+Y_2}(X) \approx \mathrm{Ch}_{Y_1}(X) + \mathrm{Ch}_{Y_2}(X).$$

Our object here is three-fold. (a) we point out how the products with G and H are defined in terms of the former products. (b) The products are discussed, assuming merely that they satisfy the (P_i) and the (Q_i). (c) In a simplicial complex, for certain groups G and H, when one of $\smile$, $\frown$ is known in the cohomology and homology groups, the other may be determined at once.

(a) If the products with integral coefficients are given, we may set

$$(\sum h_i \sigma_i^p) \cdot (\sum g_i \sigma_i^p) = \sum h_i \cdot g_i, \tag{11.2}$$

$$h\sigma_i^p \smile h'\sigma_j^q = \sum_k {}^{pq}\Gamma_k^{ij} hh' \sigma_k^{p+q}, \tag{11.3}$$

$$h\sigma_j^q \frown g\sigma_k^{p+q} = \sum_l {}^{pq}\Gamma_k^{lj} h \cdot g \sigma_l^p. \tag{11.4}$$

The relations (P_2) and (Q_2), and hence (5.6) through (5.9), continue to hold, so the products are defined among the cohomology and homology groups. If u^p and v^p denote homology and cohomology classes, the explicit definitions are

[20] As $h \cdot (g + g') = h \cdot g + h \cdot g'$, if we consider H as an additive group, then H and G form a "group pair" with respect to G. We could replace H by any subring. Thus, in (γ) below, we could take $H \approx I_0$.

$$v^p A^p \cdot u^p B^p = A^p \cdot B^p, \tag{11.5}$$

$$v^p A^p \smile v^q B^q = v^{p+q}(A^p \smile B^q), \tag{11.6}$$

$$v^q B^q \frown u^{p+q} C^{p+q} = u^p (B^q \frown C^{p+q}). \tag{11.7}$$

(5.4) holds, because of (11.1), and hence

$$(v_i^p \smile v_j^q) \cdot u_k^{p+q} = v_i^p \cdot (v_j^q \frown u_k^{p+q}). \tag{11.8}$$

(P_3) and (Q_3) with $\gamma = 1$ are replaced by the relations

$$hI \smile h'\sigma_j^q = hh'\sigma_j^q, \qquad I \cdot (h\sigma_j^q \frown g\sigma_j^q) = h \cdot g. \tag{11.9}$$

(We could leave out the h in the first.) Thus the integer $\gamma(\smile) = \gamma(\frown)$ is replaced by the operations $\gamma(\smile; h, h') = hh'$, $\gamma(\frown; h, g) = h \cdot g$; if G is the group of integers, then $\gamma = \gamma(\smile; 1, 1) = \gamma(\frown; 1, 1)$. To a general γ correspond general bilinear maps $\gamma(\smile; h, h')$ into H and $\gamma(\frown; h, g)$ into G, and (11.9) becomes

$$hI \smile h'\sigma_j^q = \gamma(\smile; h, h')\sigma_j^q, \qquad I \cdot (h\sigma_j^q \frown g\sigma_j^q) = \gamma(\frown; h, g). \tag{11.10}$$

The relations (5.10) through (5.14) and (5.4) hold, if in (5.12) we assume that H is commutative. For the proofs, we need merely multiply by elements of H and G and sum, using (11.1). The same remark holds for Theorem 2. Suppose ${}^p\partial_i^j = \pm 1$ or 0 in the complex. With the above interpretation for γ, the proofs of Theorems 1 and 3 hold. (Note that only the revised (P_i) and (Q_i) are used here.)

(b) Suppose $\smile$ and $\frown$ products are defined, satisfying the analogues of (P_1) and (Q_1), (P_2) and (Q_2), and (11.10), which we call (P_3') and (Q_3'). That $\smile$ and $\frown$ are bilinear means that there are bilinear maps ${}^{pq}\Phi_k^{ij}$ of (H, G) into G and ${}^{pq}\Psi_k^{ij}$ of (H, H) into H such that

$$h\sigma_i^p \smile h'\sigma_j^q = \sum_k {}^{pq}\Psi_k^{ij}(h, h')\sigma_k^{p+q}, \tag{11.11}$$

$$h\sigma_j^q \frown g\sigma_k^{p+q} = \sum_l {}^{pq}\Phi_k^{lj}(h, g)\sigma_l^p. \tag{11.12}$$

(P_2) and (Q_2) translate into (Γ_2) with ${}^{pq}\Gamma_k^{ij}$ replaced by ${}^{pq}\Psi_k^{ij}(h, h')$ and ${}^{pq}\Phi_k^{ij}(h, g)$ respectively. We suppose scalar products are defined by (11.2). We say $\smile$ and $\frown$ *correspond* if (5.4) holds. By (11.11) and (11.12), this is so if and only if

$${}^{pq}\Psi_k^{ij}(h, h') \cdot g = h \cdot {}^{pq}\Phi_k^{ij}(h', g). \tag{11.13}$$

(In the simplest case, $\Psi = \gamma hh'$ and $\Phi = \gamma h' \cdot g$ for some integer γ, and (11.13) follows from (11.1).) The products may not be derivable from the products in (a). For instance, if G is the group of real numbers, the ${}^{pq}\Gamma_k^{ij}$ may not be integers.

If $\frown$ *is given, and* $H = \mathrm{Ch}_G(G)$, *then* (11.13) *determines* $\smile$. For, given h and h', the right hand side is a homomorphism of G into itself, and thus deter-

mines an element of H; we call this element ${}^{pq}\Psi_k^{ij}(h, h')$. Clearly Ψ is bilinear, and (11.13) holds. If we take (Γ_2) with $\Phi(h', g)$ and apply h, using (11.13) gives a relation with terms acting on g. It being true for all g, we can drop it, and find (Γ_2) with $\Psi(h, h')$, i.e. (P_2). (Q_1) gives (P_1). By (Q_3'), i.e. the second relation in (11.10), and (5.4) (which follows from (11.13)),

$$(hI \smile h'\sigma_j^q)\cdot g\sigma_j^q = hI\cdot(h'\sigma_j^q \frown g\sigma_j^q) = h\cdot\gamma(\frown; h', g).$$

For fixed h and h', the last term is a map of G into itself, defining an element of H, which we call $\gamma(\smile; h, h')$; thus

$$\gamma(\smile; h, h')\cdot g = h\cdot\gamma(\frown; h', g). \tag{11.14}$$

Clearly $\gamma(\smile; h, h')$ is bilinear. If h^* is the coefficient of $hI \smile h'\sigma_j^q$ in σ_j^q, then the above relations give

$$h^*\cdot g = h^*\sigma_j^q\cdot g\sigma_j^q = \gamma(\smile; h, h')\cdot g, \qquad h^* = \gamma(\smile; h, h').$$

Hence, by (Q_1), the first relation in (11.10), i.e. (P_3'), holds.

Let 1 be the identity in H: $1\cdot g = g$ (all g). Consider the following *hypothesis* (H) on H, or on any subring containing 1: The only homomorphism ϕ of H into G (i.e. $\phi(h_1 + h_2) = \phi(h_1) + \phi(h_2)$) such that $\phi(1) = 0$ is $\phi(h) = 0$ (all h). This holds for instance if H is any of the rings given above. If $G = I_0 + I_0 = I_0^2$, we may use, in place of $H \approx I_0^4$, a subring $H' \approx I_0$.

If $\smile$ *is given, and* H *satisfies* (H), *then* $\frown$ *is determined.* For fixed h' and g, the left hand side of (11.13) is a homomorphism ϕ of H into G. Set $g^* = \phi(1)$. Then for all h in H, $\phi(h) = h\cdot g^*$; for if $\phi'(h) = \phi(h) - h\cdot g^*$, then $\phi'(1) = 0$. Set ${}^{pq}\Phi_k^{ij}(h', g) = g^*$; Φ is bilinear, and (11.13) holds. Applying (Γ_2) with $\Psi(h, h')$ to g and using (11.13) gives a relation for all h. Setting $h = 1$ gives (Q_2). (P_1) gives (Q_1). By (5.4) and the first relation in (11.10),

$$hI\cdot(h'\sigma_j^q \frown g\sigma_j^q) = \gamma(\smile; h, h')\sigma_j^q\cdot g\sigma_j^q = \gamma(\smile; h, h')\cdot g.$$

If we set $\gamma(\frown; h', g) = \gamma(\smile; 1, h')\cdot g$, then $\gamma(\frown; h', g)$ is bilinear and (Q_3') holds (as $1\ I\cdot g = I\cdot g$).

THEOREM 11. *Let* K *be a simplicial complex. Then with coefficient groups* G *and* $H = \mathrm{Ch}_\sigma(G)$, *any two* $\frown$ *products satisfying the new* (Q_i), *with the same* $\gamma(\frown; h, g)$, *give the same product in the cohomology and homology groups. The same is true for* $\smile$ *if* $H = \mathrm{Ch}_\sigma(G)$ *satisfies the hypothesis* (H).

Recall that (P_3') and (Q_3') are (11.10). First take a fixed vertex x_i^p in each simplex σ_i^p, and define the "join" $x_i^p\sigma_j^q$ for any σ_j^q in $\bar\sigma_i^p$ (which vanishes if σ_j^q contains x_i^p). Set $x_i^p \sum g_j\sigma_j^q = \sum g_j x_i^p\sigma_j^q$. Then if C^q is a q-G-cycle in $\bar\sigma_i^p$, $\partial(x_i^p C^q) = C^q$. (See for instance Lefschetz, [10], p. 111, (9').)

Now given $\frown_1$ and $\frown_2$ with the same γ, set $\frown = \frown_2 - \frown_1$. Then $\gamma(\frown; h, g) = 0$ (all h, g). We shall construct a bilinear product $h\sigma_i^q \wedge g\sigma_j^{p+q}$ as in Theorem 8. The proof there given holds, if we are careful to define the product for all h and g at each step. At a typical step, we have a p-G-cycle

$C^p = h\sigma_i^q \frown g\sigma_j^{p+q} + \cdots$ in $\bar{\sigma}_j^{p+q}$. To make $\partial(h\sigma_i^q \wedge g\sigma_k^{p+q}) = C^p$, we set $h\sigma_i^q \wedge g\sigma_j^{p+q} = x_j^{p+q} C^q$. Because of the new (R_2) and (R_3), the $\frown$ product vanishes in the cohomology and homology groups; it follows that $\frown_1$ and $\frown_2$ give the same products there.

To prove the last statement, we shall show that if $\gamma(\smile; h, h') = 0$, then $A \smile B \backsim 0$ for cocycles A and B. Construct $\frown$ corresponding to $\smile$; then by (11.14), $\gamma(\frown; h', g) = 0$. Hence, as we saw above, a bilinear $\wedge$ operation may be defined satisfying the revised (R_i). Next we shall construct $\vee$ so that (9.6) will hold. In terms of the coefficients defining $\wedge$ and $\vee$, this takes the form of the relation (11.13). The right hand side being given, we construct the coefficients on the left so the relation is satisfied, exactly as we constructed $\smile$ in terms of $\frown$. We can now show that $(A \smile B)\cdot g\sigma = \delta(A \vee B)\cdot g\sigma$ for cocycles A and B, as in the proof of Theorem 10. As this is true for all g, $A \smile B$ and $\delta(A \vee B)$ have the same coefficient in σ. As this is true for all σ, $A \smile B = \delta(A \vee B) \backsim 0$.

(c) For certain sets of groups G, H and Z, the homology and cohomology groups ${}^p\mathbf{H}^G$ and ${}^p\mathbf{H}_H$ satisfy

$$(11.15) \qquad {}^p\mathbf{H}^G \approx \mathrm{Ch}_Z({}^p\mathbf{H}_H), \qquad {}^p\mathbf{H}_H \approx \mathrm{Ch}_Z({}^p\mathbf{H}^G).$$

As shown in Whitney, [13], this is so whenever the following conditions are satisfied.

(1) G and H form a group pair with respect to Z.
(2) $G \approx \mathrm{Ch}_Z(H)$, $H \approx \mathrm{Ch}_Z(G)$.
(3) G and H resolve each other completely.
(4) Z is infinitely (better term: completely) divisible.

(11.15) holds for any of the examples (β), (γ), (δ) above, with $Z = G$. We can show this for (β) by first using $Z = R_1$, and noting that in the maps, only a subgroup $\approx G$ of R_1 is used.

THEOREM 12. *Let the products* (11.5) *be defined, and suppose* (11.15) *is satisfied. Then, for any* γ, *as soon as one of* $\smile$, $\frown$ *is known, the other is determined by* (11.8).

Recall that for any γ, $\smile$ and $\frown$ are uniquely defined. We shall show that, given $\smile$ or $\frown$, there is a unique corresponding $\frown$ or $\smile$ satisfying (11.8); as the correct products satisfy this relation, the theorem will be proved.

Suppose $\smile$ is defined. For fixed v_j^q and u_k^{p+q}, the left hand side of (11.8) is a homomorphism of ${}^p\mathbf{H}_H$ into Z, and hence, by (11.15), corresponds to a unique element of ${}^p\mathbf{H}^G$; we call this element $v_j^q \frown v_k^{p+q}$. Then $\frown$ is bilinear, and (11.8) is satisfied. We find $\smile$ in terms of $\frown$ in the same manner.

12. Construction of the products in low dimensional complexes

We shall construct all products $\sigma^q \frown \sigma^{p+q}$ for $p \leqq 2$ in a particular fashion; then $\sigma^p \smile \sigma^q$ is determined for $p \leqq 2$. We can then determine the $\smile$ products in the cohomology and homology groups of complexes of dimensions $\leqq 5$,

with the help of Theorem 12 and (5.12); the remaining $\frown$ products are then given with the aid of Theorem 12.

We begin by ordering the vertices of K in a fixed fashion, $x_1, x_2, \cdots$. For each σ_i^p, let V_i^p be its first vertex. For each vertex x_h of σ_i^p, let $W_i^p x_h$ be a 1-chain in $\bar{\sigma}_i^p$ whose boundary is $x_h - V_i^p$. Set (using $\gamma = 1$)

$$\sigma_i^p \frown \sigma_i^p = V_i^p, \tag{12.1}$$

$$\sigma_j^{p-1} \frown \sigma_i^p = {}^p\partial_i^j W_i^p V_j^{p-1} \qquad (\sigma_j^{p-1} \text{ in } \bar{\sigma}_i^p). \tag{12.2}$$

Finally, take any σ_j^{p-2} of $\bar{\sigma}_i^p$; say

$$\partial\sigma_i^p = \sum \alpha_k \sigma_k^{p-1}, \qquad \delta\sigma_j^{p-2} = \sum \beta_k \sigma_k^{p-1} \text{ in } \bar{\sigma}_i^p.$$

As $\partial\partial\sigma_i^p = 0$, $\sum \alpha_k\beta_k = 0$. Now (see §8)

$$C^1 = \delta\sigma_j^{p-2} \frown \sigma_i^p + \sigma_j^{p-2} \frown \partial\sigma_i^p = \sum \beta_k W_i^p V_k^{p-1} + \sum \alpha_k W_k^{p-1} V_j^{p-2}$$

is a cycle in $\bar{\sigma}_i^p$, and we may choose a 2-chain $\sigma_j^{p-2} \frown \sigma_i^p$ bounded by it.

In the simplest case, we will have

$$\partial\sigma_i^p = \sigma_s^{p-1} + \sigma_t^{p-1} + \cdots, \qquad \delta\sigma_j^{p-2} = \sigma_s^{p-1} - \sigma_t^{p-1} + \cdots,$$

and

$$C^1 = W_i^p V_s^{p-1} - W_i^p V_t^{p-1} + W_s^{p-1} V_j^{p-2} + W_t^{p-1} V_j^{p-2},$$

and we can find a possible $\sigma_j^{p-2} \frown \sigma_i^p$ at once. For most σ_j^{p-2} on σ_i^p (for instance, for all in which $V_j^{p-2} = V_s^{p-1} = V_t^{p-1}$), C^1 will vanish; but there will in general be at least one for which it does not.

III. Invariance Theorems

13. Subdivision and consolidation

Let K' be a complex, satisfying (K_1), (K_2) and (K_3) of §2. Let $\{E_i^p\}$ be a set of distinct closed subcomplexes of K' which cover K', E_i^p being of dimension p. Let $F_i^p = F(E_i^p)$ be the union of all E_j^q $(q < p)$ contained in E_i^p. Set $O_i^p = E_i^p - F_i^p$. Assume

(K_1') The common part $E_i^p \cdot E_j^q$ of any two of the subcomplexes is either void or the union of a subset of the subcomplexes.

(K_2') If E_j^q is in E_i^p, $E_j^q \neq E_i^p$, then $q < p$.

(K_3') With integer coefficients, O_i^p is monocyclic or acyclic[21] in the dimension p and is acyclic in all lower dimensions.

---

[21] O_i^p is *acyclic* in the dimension q if every q-cycle (which need not be boundary-like) is a boundary; O_i^p is *monocyclic* in the dimension p if there is a p-cycle $X^p \neq 0$ such that any p-cycle is a multiple of X^p. (All this is with integer coefficients.) Compare Tucker: "cell-like," "null-like." In [4], Ch. VI, a similar assumption, using any coefficient group G, is made. Note that O_i^p is a subcomplex of K; a chain A in O_i^p is a cycle if, considered as a chain in K, ∂A^p has no part in O_i^p.

Define a complex K as follows. With each O_i^p which is monocyclic we associate a p-cell σ_i^p. Say σ_j^q is a face of σ_i^p if and only if E_j^q is in E_i^p. We shall define the ${}^p\partial_j^i$ in K and prove (K_1), (K_2) and (K_3) for K below. We call K a *consolidation*[22] of K', and K', a *subdivision* of K. If K may be formed from a simplicial K' in this manner, we say K *admits a simplicial subdivision.*

14. Structure of K and K'

Denote the cells of K' by τ_i^p. Let J^p be the union of all E_i^q for $q \leqq p$. We prove:

(α_1) O_i^p contains all the p-cells of E_i^p; for F_i^p is of dimension $<p$.

(α_2) $E_i^p \cdot E_j^p$ is in J^{p-1} if $j \neq i$; for the common part is a union of E_k^q, and as each such E_k^q is in E_i^p, $q < p$, by (K_2').

(α_3) Each τ_i^p is in a unique O_j^q; then E_j^q is the smallest subcomplex containing τ_i^p. To prove this, choose E_j^q as stated. E_j^q is uniquely determined. For if τ_i^p is also in E_k^r, then it is in their common part and hence in an E_l^s contained in both, by (K_1'); hence $E_l^s = E_j^q$, and E_j^q is in E_k^r. As τ_i^p is in no E_k^{q-1}, it is in O_j^q. Suppose τ_i^p were also in $O_k^r \neq O_j^q$. Then E_k^r contains E_j^q, and by (K_2'), $r > q$. Hence E_j^q is in F_k^r, and τ_i^p is not in O_k^r, a contradiction.

In (α_3), we define $E_j^q = E(\tau_i^p)$, $O_j^q = O(\tau_i^p)$. If O_j^q is monocyclic (in the dimension q), we set $\sigma_j^q = \sigma(\tau_i^p)$.

(α_4) If τ is in O_i^p and in E_j^q, then E_j^q contains E_i^p, by (α_3).

For each O_i^p which is monocyclic, let

$$X_i^p = \sum_j {}^p a_i^j \tau_j^p \tag{14.1}$$

be the corresponding p-cycle; for other O_i^p, set $X_i^p = 0$. Map $L^p(K)$ into $L^p(K')$ by

$$Sd\sigma_i^p = X_i^p. \tag{14.2}$$

Define $Sd \sum \alpha_i \sigma_i^p$ by linearity. Then

(α_5) $SdA^p = 0$ implies $A^p = 0$. For if $A^p = \sum \alpha_i \sigma_i^p$, then $\sum \alpha_i X_i^p = 0$, and (α_2) shows that each $\alpha_i X_i^p = 0$; as $X_i^p \neq 0$ in this case, $\alpha_i = 0$.

LEMMA 4. *Any cycle A'^p in J^p (with integer coefficients) is SdA^p for a uniquely defined A^p in K.*

First, write $A'^p = \sum A_i'^p$, $A_i'^p$ in E_i^p. As $A'^p - A_i'^p$ is in $\sum_{j \neq i} E_j^p$, by (α_2), so is $-\partial(A'^p - A_i'^p) = \partial A_i'^p$. By (α_4), $\partial A_i'^p$ has no part in O_i^p, that is (using (α_1)), $A_i'^p$ is a p-cycle in O_i^p. Hence, if $A_i'^p \neq 0$, then O_i^p is monocyclic, and for some α_i,

$$A_i'^p = \alpha_i X_i^p; \qquad \text{then } A'^p = Sd \sum \alpha_i \sigma_i^p.$$

The uniqueness follows from (α_5). We remark that "in J^p" may be replaced by "in $J^p - J^{p-1}$."

---

[22] "*Zellenzerspaltung*" in [4], Ch. VI. But see the last foot-note.

We now define the boundary relations in K. Given σ_i^p, $\partial Sd\sigma_i^p$ is a cycle in F_i^p, which is in J^{p-1}; hence it may be written uniquely as SdB^{p-1}. *We define* $\partial\sigma_i^p = B^{p-1}$. Then

$$\partial SdA^p = Sd\partial A^p. \tag{14.3}$$

Clearly (K_1) holds for K. Also $\partial Sd\sigma_i^p$ is in E_i^p, hence $\partial\sigma_i^p$ is in $\bar{\sigma}_i^p$, and (K_2) holds. For any σ_i^p, $Sd\partial\partial\sigma_i^p = \partial\partial Sd\sigma_i^p = 0$; hence, by ($\alpha_5$), $\partial\partial\sigma_i^p = 0$, and ($K_3$) holds.

THEOREM 13. *If K' is a subdivision of K, then there are homomorphisms ϕ of $L^p(K')$ into $L^p(K)$ and ψ of $L^p(K')$ into $L^{p+1}(K')$ such that*

(a) *$\phi\tau^p$ is in $\bar{\sigma}(\tau^p)$; $\psi\tau^p$ is in $E(\tau^p)$.*

(b) $\partial\phi A'^p = \phi\partial A'^p$.

(c) $\phi SdA^p = A^p$.

(d) $Sd\phi A'^p = A'^p - \psi\partial A'^p - \partial\psi A'^p$.

If we take duals of these relations and (14.3), using:

$\phi' = D(\phi)$, mapping $L^p(K)$ into $L^p(K')$,

$Sd' = D(Sd)$, mapping $L^p(K')$ into $L^p(K)$,

$\psi' = D(\psi)$, mapping $L^{p+1}(K')$ into $L^p(K')$,

we obtain

(b') $\delta\phi' A^p = \phi'\delta A^p$,

(c') $Sd'\phi' A^p = A^p$,

(d') $\phi' Sd' A'^p = A'^p - \psi'\delta A'^p - \delta\psi' A'^p$,

$$\delta Sd' A'^p = Sd'\delta A'^p. \tag{14.3'}$$

We begin by constructing homomorphisms θ_p and ψ_p as follows. θ_p and ψ_p are defined in J^p, and map $L^r(K')$ into $L^r(K')$ and into $L^{r+1}(K')$ respectively. For τ^q in J^{p-1}, set $\theta_p\tau^q = \tau^q$, $\psi_p\tau^q = 0$. If τ^q is in O_i^p, then we shall have:

$$\psi_p\tau^q \text{ is in } O_i^p, \qquad \theta_p\tau^q \text{ is in } \begin{cases} F_i^p \text{ if } q < p, \\ E_i^p \text{ if } q = p. \end{cases}$$

Also

$$\partial\psi_p\tau^q = \tau^q - \psi_p\partial\tau^q - \theta_p\tau^q. \tag{14.4}$$

For τ^0 in J^0, set $\theta_0\tau^0 = \tau^0$, $\psi_0\tau^0 = 0$. Suppose all θ_r and ψ_r are constructed in J^{p-1}; we shall construct them in J^p. We need merely consider θ_p and ψ_p; for (14.4) holds for θ_r and ψ_r, $r > p$, by their definitions. For τ^0 in O_i^p, we may choose $\psi_p\tau^0$ in O_i^p and $\theta_p\tau^0$ in F_i^p so that

$$\partial\psi_p\tau^0 = \tau^0 - \theta_p\tau^0,$$

by (K_3'); then (14.4) holds.

Now suppose ψ_p and θ_p are defined for all cells of dimension $< q$ in O_i^p; we shall define them for τ^q. If $\partial_1\tau^q$ is the part of $\partial\tau^q$ in O_i^p, $\psi_p\partial_1\tau^q$ is in O_i^p; if $\partial_2\tau^q$ is the part in F_i^p, $\psi_p\partial_2\tau^q = 0$. Hence $\psi_p\partial\tau^q$ is in O_i^p. Also, using (14.4) applied to $\partial\tau^q$,

$$\partial(\tau^q - \psi_p\partial\tau^q) = \partial\tau^q - (\partial\tau^q - \theta_p\partial\tau^q) = \theta_p\partial\tau^q, \tag{14.5}$$

which is in F_i^p. Hence $\tau^q - \psi_p\partial\tau^q$, as a chain in the complex O_i^p, is a cycle.

Suppose first that $q < p$. Then, by (K_3'), we may find a chain $\psi_p\tau^q$ in O_i^p and a chain $\theta_p\tau^q$ in F_i^p so that (14.4) holds. Suppose next that $q = p$. Then if $\tau^q = \tau_j^p$,

$$\tau_j^p - \psi_p\partial\tau_j^p = \alpha_j X_i^p, \tag{14.6}$$

by (K_3'). If $X_i^p = 0$, we take $\alpha_j = 0$. Set $\psi_p\tau_j^p = 0$, $\theta_p\tau_j^p = \alpha_j X_i^p$; again (14.4) follows.

We prove some properties of ψ_p and θ_p. Taking the boundary of (14.4) gives

$$\begin{aligned} 0 &= \partial\tau^q - (\partial\tau^q - \psi_p\partial\partial\tau^q - \theta_p\partial\tau^q) - \partial\theta_p\tau^q, \\ \partial\theta_p\tau^q &= \theta_p\partial\tau^q. \end{aligned} \tag{14.7}$$

Next, by (14.1) and (14.6), summing over all p-cells in O_i^p,

$$\textstyle\sum_j \alpha_j {}^pa_i^j X_i^p = \sum_j {}^pa_i^j(\tau_j^p - \psi_p\partial\tau_j^p) = X_i^p - \psi_p\partial X_i^p.$$

But ∂X_i^p is in F_i^p, and hence $\psi_p\partial X_i^p = 0$. Therefore

$$\textstyle\sum_j \alpha_j {}^pa_i^j = 1 \qquad \text{if } X_i^p \neq 0. \tag{14.8}$$

It follows that, even if $X_i^p = 0$,

$$\theta_p X_i^p = \textstyle\sum_j {}^pa_i^j\theta_p\tau_j^p = \sum_j {}^pa_i^j\alpha_j X_i^p = X_i^p. \tag{14.9}$$

Now define a homomorphism θ of $L^q(K')$ into $L^q(J^q)$ by

$$\theta\tau^q = \theta_q\theta_{q+1}\cdots\theta_n\tau^q, \tag{14.10}$$

supposing K' is of dimension n. By the definition of $\theta_p\tau^q$ for $p = q$, we may define ϕ in $L^q(K')$ so that

$$\theta\tau^q = Sd\phi\tau^q. \tag{14.11}$$

By (14.3), (14.9) and (14.7),

$$\partial\theta\tau^q = Sd\partial\phi\tau^q = \theta_{q-1}Sd\partial\phi\tau^q = \theta_{q-1}\partial\theta\tau^q = \theta_{q-1}\theta_q\cdots\theta_n\partial\tau^q = \theta\partial\tau^q; \tag{14.12}$$

hence

$$Sd\partial\phi\tau^q = \partial Sd\phi\tau^q = \partial\theta\tau^q = Sd\phi\partial\tau^q,$$

and (b) follows from (α_5).

Finally, define ψ by

$$\psi\tau^q = \textstyle\sum_{k=q}^n \psi_k\theta_{k+1}\cdots\theta_n\tau^q, \tag{14.13}$$

using $\theta_{n+1} \cdots \theta_n \tau^q = \tau^q$. As $k+1 > q$ in each term of the sum, $\theta_{k+1} \cdots \theta_n \tau^q$ is in J^k, so that $\psi_k \theta_{k+1} \cdots \theta_n \tau^q$ is defined. Note that $\psi_{q-1}\theta_q \cdots \theta_n \partial\tau^q = 0$, as $\theta_q \cdots \theta_n \partial\tau^q$ is of dimension $q-1$. Hence, by (14.4) and (14.7),

$$\begin{aligned}\partial\psi\tau^q &= \textstyle\sum_{k=q}^n \partial\psi_k\theta_{k+1} \cdots \theta_n\tau^q \\ &= \textstyle\sum_{k=q}^n \theta_{k+1} \cdots \theta_n\tau^q - \sum_{k=q}^n \psi_k\partial\theta_{k+1} \cdots \theta_n\tau^q \\ &\quad - \textstyle\sum_{k=q}^n \theta_k\theta_{k+1} \cdots \theta_n\tau^q = \tau^q - \psi\partial\tau^q - \theta\tau^q;\end{aligned}$$

this, with (14.11), gives (d).

To prove (c), (14.9) gives

$$Sd\phi Sd\sigma_i^p = Sd\phi X_i^p = \theta X_i^p = \theta_p X_i^p = X_i^p = Sd\sigma_i^p;$$

(c) now follows from (α_5).

15. Combinatorial invariance

We can now prove (compare Theorem 13)

THEOREM 14. *Let K' be a subdivision of K, and let G be an abelian group. Then both ϕ and Sd induce isomorphisms between ${}^p\mathbf{H}^G(K)$ and ${}^p\mathbf{H}^G(K')$, and both ϕ' and Sd' induce isomorphisms between ${}^p\mathbf{H}_G(K)$ and ${}^p\mathbf{H}_G(K')$. If K and K' each admits a product theory, then these isomorphisms preserve products.*[23]

The meaning of the last phrase is seen from (15.4), (15.5), (15.7) and (15.8) below.

Set $Sd \sum g_i\sigma_i^p = \sum g_i Sd\sigma_i^p$ etc. Using u and v for homology and cohomology classes as in §11, set

$$\phi u' A'^p = u\phi A'^p, \qquad SduA^p = u' SdA^p, \tag{15.1}$$

$$\phi' vA^p = v'\phi' A^p, \qquad Sd'v'A'^p = vSd'A'^p, \tag{15.2}$$

the chains being cycles in (15.1) and cocycles in (15.2). The proof that these are isomorphisms follows at once from Theorem 13. Consider for instance ϕ'. To show that $\phi' v$ is uniquely determined, suppose $vA^p = vB^p$. Then $A^p - B^p = \delta C^{p-1}$, and

$$v'\phi' A^p - v'\phi' B^p = v'\phi'\delta C^{p-1} = v'\delta\phi' C^{p-1} = 0.$$

Suppose $\phi' vA^p = \phi' vB^p$. Then $\phi' A^p - \phi' B^p = \delta C'^{p-1}$, and

$$A^p - B^p = Sd'\phi'(A^p - B^p) = Sd'\delta C'^{p-1} = \delta Sd' C'^{p-1},$$

so that $vA^p = vB^p$. Given a $v'A'^p$, to find a vA^p mapping into it, set $A^p = Sd'A'^p$. Then $\delta A^p = Sd'\delta A'^p = 0$, and

$$\phi' vA^p = v'\phi' Sd' A'^p = v'(A'^p - \psi'\delta A'^p - \delta\psi' A'^p) = v'A'^p.$$

---

[23] If topological coefficient groups are used, the isomorphisms are continuous. The part of this theorem relating to the homology groups has been proved by Tucker, [12], for integral coefficients, and by Alexandroff-Hopf, [4], Ch. VI, using the stronger condition noted in footnote 21.

To prove that the $\smile$ products agree in the cohomology groups of K and K' (if they are defined), consider first integer coefficients. Take a definite $\smile$ product in K', and define a new one in K by

$$A \smile B = Sd'(\phi' A \smile \phi' B). \tag{15.3}$$

We must prove (P_1) through (P_3). To prove (P_1), say

$$\sigma_i^p \smile \sigma_j^q = \alpha \sigma_k^r + \cdots, \qquad \alpha \neq 0 \qquad (r = p + q).$$

Then

$$\phi'\sigma_i^p \smile \phi'\sigma_j^q = \beta\tau^r + \cdots, \qquad Sd'\tau^r = \gamma\sigma_k^r, \qquad \beta \neq 0, \gamma \neq 0.$$

As then $Sd\sigma_k^r = \gamma\tau^r + \cdots$, τ^r is in O_k^r, and $\sigma(\tau^r) = \sigma_k^r$. Say

$$\phi'\sigma_i^p = \epsilon\tau^p + \cdots, \qquad \phi'\sigma_j^q = \zeta\tau^q + \cdots, \qquad \tau^p \smile \tau^q = \eta\tau^r + \cdots$$

$$(\epsilon, \zeta, \eta \neq 0).$$

As then $\phi\tau^p = \epsilon\sigma_i^p + \cdots$, σ_i^p is in $\bar{\sigma}(\tau^p)$. As τ^p is in $\bar{\tau}^r$, $E(\tau^p)$ is in $E(\tau^r)$, and σ_i^p is in $\bar{\sigma}(\tau^r) = \bar{\sigma}_k^r$. Similarly σ_j^q is in $\bar{\sigma}_k^r$.

To prove (P_2), (b′) and (14.3′) with (P_2) for $\smile$ in K' give

$$\delta(A \smile B) = Sd'(\phi'\delta A \smile \phi' B \pm \phi' A \smile \phi'\delta B) = \delta A \smile B \pm A \smile \delta B.$$

To prove (P_3), note that $\phi' I = I'$; hence, using (P_3) in K',

$$I \smile \sigma_j^q = Sd'(\phi' I \smile \phi'\sigma_j^q) = Sd'\phi'\sigma_j^q = \sigma_j^q.$$

Therefore we may use the product of (15.3) in K. To prove that ϕ' and Sd' preserve products in the cohomology groups, note that (for cocycles), using (d′) and (5.6), (5.7),

$$\phi'(A \smile B) = \phi' Sd'(\phi' A \smile \phi' B) \backsim \phi' A \smile \phi' B, \tag{15.4}$$

$$Sd'A' \smile Sd'B' = Sd'(\phi' Sd'A' \smile \phi' Sd'B') \backsim Sd'(A' \smile B'). \tag{15.5}$$

We may similarly define $\frown$ in K in terms of $\frown$ in K' by

$$A \frown B = \phi(\phi' A \frown SdB). \tag{15.6}$$

This corresponds to the $\smile$ product in K, and hence is a $\frown$ product. To show this, (4.6) gives

$$(A \smile B)\cdot C = Sd'(\phi' A \smile \phi' B)\cdot C = (\phi' A \smile \phi' B)\cdot SdC,$$

$$A \cdot (B \frown C) = A \cdot \phi(\phi' B \frown SdC) = \phi' A \cdot (\phi' B \frown SdC).$$

Applying (5.4) in K' shows that these are equal; hence $\smile$ and $\frown$ correspond in K, by (5.4).

We map cycles [cocycles] from K into K' and from K' into K with Sd and ϕ [with ϕ' and Sd']. The invariance of the $\frown$ product is given by

$$Sd(A \frown B) = Sd\phi(\phi' A \frown SdB) \sim \phi' A \frown SdB, \tag{15.7}$$

$$Sd'A' \frown \phi B' = \phi(\phi' Sd'A' \frown Sd\phi B') \sim \phi(A' \frown B'), \tag{15.8}$$

where A, A' are cocycles, B, B' are cycles.

Now consider the coefficient group G as in §11, (a). The $\smile$ products in the cohomology groups in both K and K' are formed by (11.3). As (15.3) holds with A and B replaced by ${}^{p}h\sigma_i^p$ and ${}^{q}h\sigma_j^q$, (15.3) etc. hold with any coefficient group. (15.4) and (15.5) prove the invariance. Similarly for the $\frown$ product.

16. Topological invariance

We shall show how to associate homology and cohomology groups and a product theory with a polyhedron P by means of any simplicial subdivision. By Theorem 14 we may find these groups and products, using any complex K which admits, as a subdivision, a simplicial triangulation of P, and admits a product theory.

The theorem and proof extend at once to prove the existence of groups and products in a bicompact space; compare Steenrod (see footnote 8), §9.

The proof is based on Theorem 13. However, if we restrict ourselves to simplicial complexes, Theorem 13 becomes much more simple. Hence we give it again for this case, as a lemma. We may then prove, as before, the first part of Theorem 14.

LEMMA 6. *Let K' be a simplicial subdivision of the simplicial complex K, and let ϕ be a pseudo-identical map*[24] *of K' into K. Then there is a map ψ as in Theorem* 13 *such that the conclusions of Theorem* 13 *hold.*

The statements about ϕ and Sd are well-known; we shall construct ψ. For each vertex x' of K', let $\psi x'$ be any 1-chain in the subdivision of the smallest cell $\sigma(x')$ of K containing x', which is bounded by $x' - \phi x'$. Suppose ψ is constructed in $L^0(K'), \cdots, L^{p-1}(K')$. Then applying (d) of Theorem 13 to the $(p-1)$-chain $\partial\tau^p$, we find

$$\partial(Sd\phi\tau^p - \tau^p + \psi\partial\tau^p) = \partial Sd\phi\tau^p - \partial\tau^p + \partial\tau^p - Sd\phi\partial\tau^p = 0.$$

Hence $Sd\phi\tau^p - \tau^p + \psi\partial\tau^p$ is a p-cycle in the subdivision of $\sigma(\tau^p)$ $(p > 0)$, and therefore, as is well-known, we can find a chain $-\psi\tau^p$ there bounded by it. The ψ as thus constructed clearly has the required properties.

THEOREM 15. *Let K and K' be simplicial triangulations of homeomorphic polyhedra P and P'. Then there are isomorphisms between ${}^{p}\mathbf{H}^G(K)$ and ${}^{p}\mathbf{H}^G(K')$ and between ${}^{p}\mathbf{H}_G(K)$ and ${}^{p}\mathbf{H}_G(K')$ which preserve the $\smile$ and $\frown$ products.*

---

[24] That is, a simplicial map such that each vertex x' of K' goes into a vertex of a cell of K containing it.

By identifying corresponding points of P and P', we may suppose K and K' are triangulations of P. We may subdivide K' into K_1', then K into K_1, and then K_1' into K_2', so that the following conditions hold.[25] There are simplicial maps ψ_2, ϕ_1, ψ_1, of K_2' into K_1, K_1 into K_1', and K_1' into K, such that $\Psi = \phi_1\psi_2$ and $\Phi = \psi_1\phi_1$ are pseudo-identical. The duals ψ_1', ϕ_1', ψ_2' of ψ_1, ϕ_1, ψ_2 induce homomorphisms of the cohomology groups of K, K_1', K_1 into those of K_1', K_1, K_2'. Combining pairs of these homomorphisms gives homomorphisms of ${}^p\mathbf{H}_G(K)$ into ${}^p\mathbf{H}_G(K_1)$ and of ${}^p\mathbf{H}_G(K_1')$ into ${}^p\mathbf{H}_G(K_2')$. These are induced by the duals Φ' and Ψ' of Φ and Ψ, and are isomorphisms, by Theorem 14.[26] Set $\psi^* = (\Phi')^{-1}\phi_1'$; then ψ_1' and ψ^* map ${}^p\mathbf{H}_G(K)$ into ${}^p\mathbf{H}_G(K_1')$ and vice versa. Further, using E for the identity, we have, in the cohomology groups,

$$\begin{aligned}\psi^*\psi_1' &= (\phi_1'\psi_1')^{-1}\phi_1'\psi_1' = E,\\ \psi_1'\psi^* &= (\Psi'^{-1}\Psi')\psi_1'(\phi_1'\psi_1')^{-1}\phi_1'\\ &= (\psi_2'\phi_1')^{-1}\psi_2'(\phi_1'\psi_1')(\phi_1'\psi_1')^{-1}\phi_1' = E.\end{aligned}$$

It follows that ψ_1' and ψ^* are isomorphisms (see for instance [4], p. 558). Combining this isomorphism with an isomorphism χ between ${}^p\mathbf{H}_G(K_1')$ and ${}^p\mathbf{H}_G(K')$ (Theorem 14) gives an isomorphism θ between ${}^p\mathbf{H}_G(K)$ and ${}^p\mathbf{H}_G(K')$. By the same process it is seen that ψ_1 induces an isomorphism between ${}^p\mathbf{H}^G(K_1')$ and ${}^p\mathbf{H}^G(K)$. As ψ_1 is simplicial, the isomorphisms induced by ψ_1' and ψ_1 preserve products (see §7). The same is true of χ and its dual, and hence of θ and its dual. This completes the proof.

IV. Manifolds

17. Dual complexes in a manifold

Let K be a subdivision of a closed oriented combinatorial manifold,[27] and let K' be the first derived (simplicial) subdivision of K. Order the vertices of K' by choosing first the vertices of K, next the centers of 1-cells of K, etc. Say K is of dimension n. For each cell σ_i^p of K, let E_i^{n-p} be the subcomplex of K' containing all $(n-p)$-cells of K' which have the center of σ_i^p as their first vertex, and all faces of these cells. The hypotheses on K show that the complex K^* thus formed, the "dual" of K, admits K' as a simplicial subdivision (see §13), and admits a product theory. The maps Sd and Sd^* of $L^p(K)$ and $L^p(K^*)$ into $L^p(K')$ are defined in the natural manner; for the latter, see (19.4). Define ϕ and ϕ^* as in Part III.

---

[25] See J. W. Alexander, *Combinatorial Analysis Situs*, Trans. Am. Math. Soc., vol. 28 (1926), pp. 308-310.

[26] Hence $\Phi' = \phi_1'\psi_1'$ has an inverse; but this alone does not imply that ϕ_1' and ψ_1' have inverses.

[27] Compare Seifert-Threlfall, *Topologie*, Ch. X, or Lefschetz, [10], Ch. III.

18. The Poincaré duality theorem

Using K and K^*, we may find at once the form of the duality theorem given by Kolmogoroff, [8], and Čech, [5]. Let $\tau_i^{n-p} = \mathcal{D}(\sigma_i^p)$ be the cell of K^* dual to σ_i^p; set $\sigma_i^p = \mathcal{D}^*(\tau_i^{n-p})$. Recall that

$$\partial\mathcal{D}(\sigma^{p-1}) = (-1)^p\mathcal{D}\sigma^p + \cdots \text{ if and only if } \partial\sigma_i^p = \sigma^{p-1} + \cdots ; \tag{18.1}$$

it follows that

$$\partial\mathcal{D}(A^p) = (-1)^{p+1}\mathcal{D}(\delta A^p), \qquad \delta\mathcal{D}(A^p) = (-1)^p\mathcal{D}(\partial A^p). \tag{18.2}$$

Similar relations hold for $\mathcal{D}^*$. From (18.2) we may conclude at once that $\mathcal{D}$ establishes an isomorphism between ${}^p\mathbf{H}_G(K)$ and ${}^{n-p}\mathbf{H}^G(K^*)$. But ϕSd^* establishes an isomorphism between ${}^{n-p}\mathbf{H}^G(K^*)$ and ${}^{n-p}\mathbf{H}^G(K) = {}^{n-p}\mathbf{H}^G$ (see Theorem 13); hence

$${}^p\mathbf{H}_G \approx {}^{n-p}\mathbf{H}^G. \tag{18.3}$$

Suppose that, as in §11, (c), ${}^p\mathbf{H}_G(K) \approx \mathrm{Ch}_Z({}^p\mathbf{H}^H(K))$; then

$${}^{n-p}\mathbf{H}^G \approx \mathrm{Ch}_Z({}^p\mathbf{H}^H). \tag{18.4}$$

19. Products and intersections[28]

Supposing K *is simplicial*, we shall (a) define intersections in terms of the $\frown$ product, (b) give a relation defining $\mathcal{D}$, or rather $Sd^*\mathcal{D}$, (c) find the relation between $\smile$ and $\frown$ in the cohomology and homology groups, and (d) relate $\smile$ and intersections.

(a) The intersection of a chain A^p of K and a chain B^{*q} of K^* is the following chain[29] of K':

$$A^p \circ B^{*q} = \phi^{*\prime}\mathcal{D}A^p \frown Sd^*B^{*q} \quad (\frown \text{ product from §6}). \tag{19.1}$$

We may deduce the ordinary boundary relation:

$$\begin{aligned} \partial(A^p \circ B^{*q}) \cdot &= (-1)^{q-(n-p)}\delta\phi^{*\prime}\mathcal{D}A^p \frown Sd^*B^{*q} + \phi^{*\prime}\mathcal{D}A^p \frown \partial Sd^*B^{*q} \\ &= (-1)^{n-q}\phi^{*\prime}\mathcal{D}\partial A^p \frown Sd^*B^{*q} + \phi^{*\prime}\mathcal{D}A^p \frown Sd^*\partial B^{*q} \\ &= (-1)^{n-q}\partial A^p \circ B^{*q} + A^p \circ \partial B^{*q}. \end{aligned} \tag{19.2}$$

Note that, by (7.2) and Theorem 13, (c),

$$\phi^*(A^p \circ B^{*q}) = \mathcal{D}A^p \frown B^{*q}. \tag{19.3}$$

---

[28] Compare Čech, [5]; Freudenthal, [6].

[29] We wish $A \circ B^*$ to be a cycle if A and B^* are cycles. To apply the $\smile$ or $\frown$ product, we must turn at least one of them into a cocycle; we use $\mathcal{D}A$. It is best to use a fixed $\frown$ product, which we may do in K'. We map a cocycle of K^* into a cocycle of K' with $\phi^{*\prime}$, and a cycle of K^* into a cycle of K' with Sd^*. In this manner the form of (19.1) is determined.

(b) Order the vertices of K' in the *opposite* manner from that used above. Define ϕ by mapping each vertex of K' into the last vertex of the simplex of K containing it. If Z^n is the fundamental n-cycle of K, then $Z'^n = SdZ^n$ is the fundamental n-cycle of K'. The map $Sd^*\mathfrak{D}$ is given by

$$Sd^*\mathfrak{D}A^p = \phi' A^p \frown Z'^n \qquad (\frown \text{ from } \S 6). \tag{19.4}$$

Further, as $\phi Z'^n = \phi SdZ^n = Z^n$, (7.2) gives

$$\phi Sd^*\mathfrak{D}A^p = A^p \frown \phi Z'^n = A^p \frown Z^n; \tag{19.5}$$

it is this map that Čech uses in place of $\mathfrak{D}$.

(c) Set $\theta A^p = A^p \frown Z^n$. Then, by (6.2),

$$\theta(A \smile B) = A \frown \theta B. \tag{19.6}$$

As Sd^* and ϕ induce isomorphisms in the homology groups, (19.5) shows that θ induces the same isomorphism of ${}^p\mathbf{H}_G$ into ${}^{n-p}\mathbf{H}^G$ that $\mathfrak{D}$ does. Hence θ^{-1} exists in these groups, and (19.6) gives

$$v_1 \smile v_2 = \theta^{-1}(v_1 \frown \theta v_2), \qquad v_1 \frown u_1 = \theta(v_1 \smile \theta^{-1}u_1). \tag{19.7}$$

(d) By (19.1), (19.4) and (6.2),

$$\mathfrak{D}^*A^* \circ \mathfrak{D}B = \phi^{*\prime}A^* \frown Sd^*\mathfrak{D}B = (\phi^{*\prime}A^* \smile \phi' B) \frown Z'^n.$$

By (c), applying this to the cohomology groups gives

$$\theta v_1 \circ \theta v_2 = \theta(v_1 \smile v_2), \tag{19.8}$$

and hence

$$u_1 \circ u_2 = \theta(\theta^{-1}u_1 \smile \theta^{-1}u_2). \tag{19.9}$$

A final remark. For a positively oriented σ^n, $\sigma^n \frown Z^n$ is the first vertex of σ^n. Hence

$$I \cdot (A^n \frown Z^n) = A^n \cdot Z^n. \tag{19.10}$$

20. On intersections of chains and complexes

Let M^n be a manifold, and let K and K' be singular complexes (i.e. continuous maps of complexes into M^n) in "general position" in M^n; that is, so that no σ^p intersects any σ'^{n-p-1}. (A slight deformation of K' will bring this about.) Then all Kronecker indices $(\sigma_i^p \circ \sigma_j'^{n-p})$ have meaning, and $(A^p \circ \partial B'^{n-p+1}) = (-1)^p(\partial A^p \circ B'^{n-p+1})$. (See for instance Lefschetz, [10], p. 169, (20).) Set

$$g\sigma_i^p = \textstyle\sum_j (\sigma_i^p \circ \sigma_j'^{n-p})\sigma_j'^{n-p}, \qquad g'\sigma_i'^{n-p} = \sum_j (\sigma_j^p \circ \sigma_i'^{n-p})\sigma_j^p, \tag{20.1}$$

and hence define gA^p, $g'A'^p$. These are dual. We may call the chain gA^p of K' *the intersection of the chain* A^p *of* K *with* K'.

For A^p in K and B'^{n-p} in K', we have the Kronecker index

$$(A^p \circ B'^{n-p}) = gA^p \cdot B'^{n-p} = A^p \cdot g'B'^{n-p}. \tag{20.2}$$

As

$$\delta gA^p \cdot \sigma' = gA^p \cdot \partial\sigma' = (A^p \circ \partial\sigma') = (-1)^p(\partial A^p \circ \sigma') = (-1)^p g\partial A^p \cdot \sigma',$$

we have $\delta g = (-1)^p g\partial$. Taking duals also, we have[30]

$$\delta gA^p = (-1)^p g\partial A^p, \qquad \delta g'A'^p = (-1)^p g'\partial A'^p. \tag{20.3}$$

In particular, cycles of K or K' map into cocycles of K' or K.

We may let K be a subdivision of M^n and let K' be a deformed position of K; then chains of K are mapped into chains of K', which may be considered as chains of K again. Then g takes the place of the $\mathfrak{D}$ in §18.

21. Dual bases

In this section, *we shall use the group R of real or of rational numbers* as coefficient group. Say a set of p-cycles $X_1^p, \cdots, X_s^p$ forms a *base* if they are linearly independent with homology, i.e.

$$\alpha_1 X_1^p + \cdots + \alpha_s X_s^p \sim 0 \quad \text{implies} \quad \alpha_1 = \cdots = \alpha_s = 0,$$

and if any p-cycle is homologous to a linear combination of them with real or rational coefficients. In other words, their homology classes form a base for $\mathbf{H}^p = {}^p\mathbf{H}^R$. Define a base for p-cocycles similarly. *Bases exist in any complex.* To show this, note first that $\mathbf{H}^p$ is a vector group; for it has a finite number of generators, (using elements of R as coefficients), and no elements of finite order. (If $kX^p \sim 0$, $k \neq 0$, then $kX^p = \partial Y^{p+1}$, $X^p = \partial(Y^{p+1}/k) \sim 0$.) Hence we may choose independent generators $u_1, \cdots, u_s$. Let X_i^p be a cycle in the class u_i; then $X_1^p, \cdots, X_s^p$ form a base. Say a base $X_1^p, \cdots, X_s^p$ for p-cycles and a base $C_1^p, \cdots, C_t^p$ for p-cocycles (then $t = s$) are *dual* if

$$C_i^p \cdot X_j^p = \delta_{ij} \quad (= 1 \text{ if } i = j, \text{ and } = 0 \text{ if } i \neq j). \tag{21.1}$$

Dual bases exist in any complex. First, let $X_1^p, \cdots, X_s^p$ be a base for p-cycles. As $\mathbf{H}_p = {}^p\mathbf{H}_R \approx \mathrm{Ch}_R\mathbf{H}^p$, the group of characters of $\mathbf{H}^p$ into R, (see for instance Whitney, [13], Theorems 7 and 8), we may choose cocycles $C_1^p, \cdots, C_s^p$ such that (21.1) holds. Let v_i be the cohomology class of C_i^p. Clearly any character of $\mathbf{H}^p$ may be expressed uniquely as a linear combination of $v_1, \cdots, v_s$; hence $C_1^p, \cdots, C_s^p$ form a base for p-cocycles.

Now consider a closed orientable manifold M^n. Dual bases for n-cycles and n-cocycles are formed by the fundamental n-cycle Z^n and a single n-cell σ^n, oriented so that $\sigma^n \cdot Z = 1$; similarly a vertex x and the cocycle I form dual

[30] The converse relations $\partial g = \pm g\delta$ etc. are false in general.

bases. Say bases $C_1^p, \cdots, C_s^p$ and $D_1^{n-p}, \cdots, D_t^{n-p}$ form *dual bases* for p-cocycles and $(n-p)$-cocycles (then $t = s$) if

$$C_i^p \smile D_j^{n-p} \frown \delta_{ij}\sigma^n, \qquad \text{i.e.} \qquad (C_i^p \smile D_j^{n-p}) \cdot Z^n = \delta_{ij}. \tag{21.2}$$

We may have $n - p = p$. Then if

$$X_i^{n-p} = C_i^p \frown Z^n, \qquad Y_i^p = D_i^{n-p} \frown Z^n, \tag{21.3}$$

we find, using (5.4) and (5.12),

$$C_i^p \cdot Y_i^p = \delta_{ij}, \qquad D_i^{n-p} \cdot X_j^{n-p} = (-1)^{p(n-p)}\delta_{ij}.$$

Hence the C_i^p and Y_i^p, also the $(-1)^{p(n-p)}D_i^{n-p}$ and X_i^{n-p}, form dual bases.

Finally, the X_i^{n-p} and Y_i^p form dual bases in the ordinary sense. For, using simple properties of intersections and (19.8) and (19.10),

$$\begin{aligned}(X_i^{n-p} \circ Y_j^p) &= I' \cdot (\theta C_i^p \circ \theta D_j^{n-p}) = I \cdot \theta(C_i^p \smile D_j^{n-p}) \\ &= I \cdot [(C_i^p \smile D_j^{n-p}) \frown Z^n] = (C_i^p \smile D_j^{n-p}) \cdot Z^n = \delta_{ij}.\end{aligned}$$

V. Products in Product Complexes

22. Definition of the products

Let K_1 and K_2 be two complexes (simplicial or not), with cells σ_i^p and τ_k. Then we have (properly oriented) $(p+r)$-cells $\sigma_i^p \times \tau_k$ and $(p+r)$-chains $A^p \times B^r$ in the product complex $K^* = K_1 \times K_2$. We recall that

$$\partial(\sigma^p \times \tau^r) = (\partial\sigma^p \times \tau^r) + (-1)^p(\sigma^p \times \partial\tau^r). \tag{22.1}$$

It follows that

$$\delta(\sigma^p \times \tau^r) = (\delta\sigma^p \times \tau^r) + (-1)^p(\sigma^p \times \delta\tau^r). \tag{22.2}$$

Choose products in K_1 and K_2. We define products in $K_1 \times K_2$ in terms of these by

$$(\sigma^p \times \tau^r) \smile (\sigma^q \times \tau^s) = (-1)^{qr}(\sigma^p \smile \sigma^q) \times (\tau^r \smile \tau^s), \tag{22.3}$$

$$(\sigma^p \times \tau^r) \frown (\sigma^q \times \tau^s) = (-1)^{p(s-r)}(\sigma^p \frown \sigma^q) \times (\tau^r \frown \tau^s). \tag{22.4}$$

We shall show that[31] the $\smile$ product has the required properties in $K_1 \times K_2$. (P_1) of §5 clearly holds. Also,

$$I^* \smile I^* = (I_1 \times I_2) \smile (I_1 \times I_2) = (I_1 \smile I_1) \times (I_2 \smile I_2) = I_1 \times I_2 = I^*,$$

[31] Of course $K_1 \times K_2$ as above defined admits a simplicial subdivision, with the proper geometric interpretation. Hence, by Theorem 14, we obtain the correct products from $K_1 \times K_2$.

so that $\gamma = 1$. Now let us calculate the terms in (P_2). We find

$$\delta[(\sigma^p \times \tau^r) \smile (\sigma^q \times \tau^s)]$$
$$= (-1)^{qr}[(\delta\sigma^p \smile \sigma^q) \times (\tau^r \smile \tau^s) + (-1)^p(\sigma^p \smile \delta\sigma^q) \times (\tau^r \smile \tau^s)]$$
$$+ (-1)^{p+q(r+1)}[(\sigma^p \smile \sigma^q) \times (\delta\tau^r \smile \tau^s) + (-1)^r(\sigma^p \smile \sigma^q) \times (\tau^r \smile \delta\tau^s)],$$
$$\delta(\sigma^p \times \tau^r) \smile (\sigma^q \times \tau^s) = (-1)^{qr}(\delta\sigma^p \smile \sigma^q) \times (\tau^r \smile \tau^s)$$
$$+ (-1)^{p+q(r+1)}(\sigma^p \smile \sigma^q) \times (\delta\tau^r \smile \tau^s),$$
$$(\sigma^p \times \tau^r) \smile \delta(\sigma^q \times \tau^s) = (-1)^{(q+1)r}(\sigma^p \smile \delta\sigma^q) \times (\tau^r \smile \tau^s)$$
$$+ (-1)^{q(r+1)}(\sigma^p \smile \sigma^q) \times (\tau^r \smile \delta\tau^s).$$

From these equations (P_2) follows at once. We may prove similarly that the $\frown$ product has the required properties, or that it corresponds to the $\smile$ product.

In the product K of three complexes, the $\smile$ and $\frown$ products come out the same whether we write K in the form $(K_1 \times K_2) \times K_3$ or $K_1 \times (K_2 \times K_3)$. The signs are as in (23.8) and (23.9) below.

23. The products in Euclidean space

We may subdivide Euclidean n-space $E^n = (u_1, \cdots, u_n)$ by means of the planes $u_i =$ an integer $(i = 1, \cdots, n)$. (We could subdivide similarly any small portion of a differentiable manifold.) We shall work out explicitly a $\smile$ product in E^n by using the product of §6 and writing $E^n = E^1 \times \cdots \times E^1$.

First consider E^1. We may denote its cells by

(23.1) $$(\alpha, \beta), \qquad \alpha \text{ an integer}, \beta = 0 \text{ or } 1;$$

it is either the vertex $u = \alpha$ or the 1-cell $\alpha \leqq u \leqq \alpha + \beta$; its dimension is β. Set $(\alpha, \beta) = 0$ for any other β. If we order the vertices by letting $(\alpha, 0)$ precede $(\alpha', 0)$ if $\alpha < \alpha'$, then the $\smile$ product of §6 may be written

(23.2) $$(\alpha, \beta) \smile (\alpha + \beta, \gamma) = (\alpha, \beta + \gamma),$$

the product vanishing in all other cases.

We turn now to the n-dimensional case. The cells of E^n are

(23.3) $$(\alpha_1, \beta_1; \cdots; \alpha_n, \beta_n), \qquad \alpha_i \text{ integral}, \beta_i = 0 \text{ or } 1;$$

the dimension of such a cell is $\sum \beta_i$. Any such symbol with some $\beta_i \neq 0$ or 1 is set $= 0$. Set

(23.4) $$S_i(\beta) = S_i(\beta_1, \cdots, \beta_n) = (-1)^{\beta_1 + \cdots + \beta_{i-1}};$$

in particular, $S_1(\beta) = 1$. The boundary relations are

(23.5) $$\partial(\alpha_1, \beta_1; \cdots; \alpha_n, \beta_n) = \textstyle\sum_i S_i(\beta)(\alpha_1, \beta_1; \cdots; \alpha_i + 1, \beta_i - 1; \cdots)$$
$$- \textstyle\sum_i S_i(\beta)(\alpha_1, \beta_1; \cdots; \alpha_i, \beta_i - 1; \cdots),$$

$$\delta(\alpha_1, \beta_1; \cdots; \alpha_n, \beta_n) = -\sum_i S_i(\beta)(\alpha_1, \beta_1; \cdots; \alpha_i, \beta_i + 1; \cdots) + \sum_i S_i(\beta)(\alpha_1, \beta_1; \cdots; \alpha_i - 1, \beta_i + 1; \cdots). \tag{23.6}$$

Set also

$$P(\beta; \gamma) = P(\beta_1, \cdots, \beta_n, \gamma_1, \cdots, \gamma_n) = (-1)^\epsilon, \qquad \epsilon = \sum_{i>j} \beta_i \gamma_j. \tag{23.7}$$

Suppose each β_i and γ_i is 0 or 1, and $\beta_i + \gamma_i \leqq 1$. Then, dropping out all zeros, $P(\beta; \gamma)$ is the sign of the permutation which carries the β's and γ's from their positions shown to their positions in $(\beta_1 + \gamma_1, \cdots, \beta_n + \gamma_n)$. From (22.3) and (23.2) we now find

$$(\alpha_1, \beta_1; \cdots; \alpha_n, \beta_n) \smile (\alpha_1 + \beta_1, \gamma_1; \cdots; \alpha_n + \beta_n, \gamma_n) = P(\beta; \gamma)(\alpha_1, \beta_1 + \gamma_1; \cdots; \alpha_n, \beta_n + \gamma_n), \tag{23.8}$$

the product vanishes in all other cases. As an example,

$$(\alpha_1, 1; \alpha_2, 0) \smile (\alpha_1 + 1, 0; \alpha_2, 1) = (\alpha_1, 1; \alpha_2, 1),$$
$$(\alpha_1, 0; \alpha_2, 1) \smile (\alpha_1, 1; \alpha_2 + 1, 0) = -(\alpha_1, 1; \alpha_2, 1).$$

By taking the dual of (23.8) as in (5.4), we find at once

$$(\alpha_1 + \gamma_1 - \beta_1, \beta_1; \cdots; \alpha_n + \gamma_n - \beta_n, \beta_n) \frown (\alpha_1, \gamma_1; \cdots; \alpha_n, \gamma_n) = P(\gamma - \beta, \beta)(\alpha_1, \gamma_1 - \beta_1; \cdots; \alpha_n, \gamma_n - \beta_n). \tag{23.9}$$

It is not hard to prove the formulas for $\delta(A \smile B)$ and $\partial(A \frown B)$ from the above formulas, and thus show directly that they are the required products.

24. On the maps of products of manifolds into a complex

We shall use only chains with real or rational coefficients; then dual bases may be defined as in §20. We begin by proving, after Hopf,

THEOREM 16.[32] *Let M^n be a closed orientable manifold with fundamental n-cycle Z^n, and let f be a simplicial map of M^n into a complex K'. Let $C_1^p, \cdots, C_s^p$ and $D_1^q, \cdots, D_s^q$ be dual bases for p-cocycles and q-cocycles in M^n ($p + q = n$; p may $= q$), and let $C_1'^p, \cdots$, and $D_1'^q, \cdots$ be bases in K'. Say*

$$f'C_i'^p \frown \sum_j \lambda_{ij} C_j^p, \qquad f'D_j'^q \frown \sum_j \mu_{ij} D_j^q. \tag{24.1}$$

If $fZ^n \sim 0$, then

$$N_{ij} = \sum_{k=1}^{s} \lambda_{ik} \mu_{jk} = 0 \qquad \text{(all } i, j\text{)}. \tag{24.2}$$

---

[32] This theorem was communicated to me by H. Hopf in a letter of September 9, 1937, together with part of Theorem 17. The results (with the complex replaced by a manifold) are corollaries of his paper [7]; we base them on the corresponding relations (7.2).

First we shall find another interpretation of the λ_{ij} and μ_{ij}. Let $\{X_i^p\}$, $\{Y_i^q\}$, $\{X_i'^p\}$, $\{Y_i'^q\}$ be bases for cycles dual to $\{C_i^p\}$, $\{D_i^q\}$, $\{C_i'^p\}$, and $\{D_i'^q\}$ respectively. Say

$$fX_i^p \sim \textstyle\sum_j \xi_{ij} X_j'^p, \qquad fY_i^q \sim \sum_j \eta_{ij} Y_j'^q; \tag{24.3}$$

then

$$C_i'^p \cdot fX_j^p = \xi_{ji} = f'C_i'^p \cdot X_j^p = \lambda_{ij}, \tag{24.4}$$

$$D_i'^q \cdot fY_j^q = \eta_{ji} = f'D_i'^q \cdot Y_j^q = \mu_{ij}, \tag{24.5}$$

so that $N_{ij} = \sum \xi_{ki}\eta_{kj}$.

To prove the theorem, we need merely note that

$$\begin{aligned} 0 &= (C_i'^p \smile D_j'^q)\cdot fZ^n = f'(C_i'^p \smile D_j'^q)\cdot Z^n = (f'C_i'^p \smile f'D_j'^q)\cdot Z^n \\ &= \sum_{k,l} \lambda_{ik}\mu_{jl}(C_k^p \smile D_l^q)\cdot Z^n = \sum_{k,l} \lambda_{ik}\mu_{jl}\delta_{kl} = N_{ij}. \end{aligned}$$

THEOREM 17. *Let M^p be a closed orientable p-manifold, and let S^q be a q-sphere, $q \geqq p$. Let $\bar{X}^p$ and $\bar{Y}^q$ be corresponding fundamental cycles. Let M^{p+q} be a simplicial subdivision of $K^{p+q} = M^p \times S^q$, with fundamental cycle Z^{p+q}. If x, y are vertices of M^p, S^q, then*

$$X^p = Sd(\bar{X}^p \times y) \quad \text{and} \quad Y^q = Sd(x \times \bar{Y}^q) \tag{24.6}$$

are cycles of M^{p+q}. Let f be a simplicial map of M^{p+q} into a complex K', and suppose $fZ^{p+q} \sim 0$.

(a) If $q > p$ or $q = p$ is even, then either $fX^p \sim 0$ or $fY^q \sim 0$.

(b) If $q = p$ is odd, then one of fX^p, fY^q is homologous to a multiple of the other.

If $q > p$, then as $\mathbf{H}^r(S^q) = 0$ $(r = 1, \cdots, q-1)$, it follows that X^p and Y^q generate $\mathbf{H}^p(M^{p+q})$ and $\mathbf{H}^q(M^{p+q})$ respectively. Let I and J be the sums of vertices, and let σ^p and τ^q be cells, in M^p and S^q. Let ϕ and Sd be the maps of Theorem 13, and set

$$C^p = \phi'(\sigma^p \times J), \qquad D^q = \phi'(I \times \tau^q), \qquad \sigma^{p+q} = \sigma^p \times \tau^q. \tag{24.7}$$

These are cocycles. Then, using Theorem 13, (c).

$$C^p \cdot X^p = (\sigma^p \times J)\cdot\phi Sd(\bar{X}^p \times y) = 1, \qquad D^q \cdot Y^q = 1, \tag{24.8}$$

so that X^p and C^p, also Y^q and D^q, form dual bases. Further, as $\bar{Z}^{p+q} = \phi Z^{p+q}$ is the fundamental cycle of K^{p+q}, (15.4) gives

$$\begin{aligned} (C^p \smile D^q)\cdot Z^{p+q} &= \phi'[(\sigma^p \times J) \smile (I \times \tau^q)]\cdot Z^{p+q} \\ &= (\sigma^p \times \tau^q)\cdot\phi Z^{p+q} = \sigma^{p+q}\cdot\bar{Z}^{p+q} = 1, \end{aligned} \tag{24.9}$$

so that C^p and D^q form dual bases. Hence, defining the $X_i'^p$ and $Y_j'^q$ as before, and defining ξ_i, η_i by

$$fX^p \sim \textstyle\sum \xi_i X_i'^p, \qquad fY^q \sim \sum \eta_i Y_i'^q, \tag{24.10}$$

Theorem 16 gives, with (24.4) and (24.5),

$$\xi_i \eta_j = 0 \qquad \text{(all } i, j\text{)}. \tag{24.11}$$

It follows that all the $\xi_i = 0$ or all the $\eta_i = 0$, so that $fX^p \sim 0$ or $fY^q \sim 0$.

Now suppose $p = q$. Then X^p and Y^p together generate $\mathbf{H}^p(M^{2p})$ and C^p and D^p together generate $\mathbf{H}_p(M^{2p})$. The relations (24.6) through (24.9) still hold. Also,

$$\begin{gathered} D^p \smile C^p \backsim (-1)^{p^2} C^p \smile D^p, \qquad (D^p \smile C^p) \cdot Z^{2p} = (-1)^p, \\ (C^p \smile C^p) \cdot Z^{2p} = (D^p \smile D^p) \cdot Z^{2p} = 0, \end{gathered} \tag{24.12}$$

so that $\{C^p, D^p\}$ and $\{D^p, (-1)^p C^p\}$ form dual bases for p-cocycles. Further,

$$C^p \cdot Y^p = (\sigma^p \times J) \cdot (x \times \bar{Y}^p) = 0, \qquad D^p \cdot X^p = 0, \tag{24.13}$$

so that $\{X^p, Y^p\}$ and $\{C^p, D^p\}$, and also $\{Y^p, (-1)^p X^p\}$ and $\{D^p, (-1)^p C^p\}$, form dual bases. Comparing (24.10) with (24.3), in which $Y_i'^p = X_i'^p$, we see that (ξ_i, η_i) takes the place of[33] $(\xi_{1i}, \xi_{2i}, \cdots)$, while $(\eta_j, (-1)^p \xi_j)$ takes the place of $(\eta_{1j}, \eta_{2j}, \cdots)$. Hence, by Theorem 16,

$$\xi_i \eta_j + (-1)^p \xi_j \eta_i = 0 \qquad \text{(all } i, j\text{)}. \tag{24.14}$$

Suppose that $p = q$ is even. If not all the ξ_i are $= 0$, say $\xi_\nu \neq 0$; then setting $i = j = \nu$ in (24.14) shows that $2\xi_\nu \eta_\nu = 0$, $\eta_\nu = 0$, and then setting $i = \nu$, any j, shows that all other $\eta_j = 0$. Thus the theorem is proved for this case.

Finally, if $p = q$ is odd, (24.14) shows that all determinants of the matrix

$$\begin{Vmatrix} \xi_1 & \xi_2 & \cdots \\ \eta_1 & \eta_2 & \cdots \end{Vmatrix}$$

vanish, so the two sets of numbers are proportional. Thus (b) of the theorem holds.

COROLLARY.[34] *If n is even and f is a map of $S^n \times S^n$ into a complex K' of dimension $< 2n$, then one of $S^n \times y$, $x \times S^n$ is mapped into a cycle ~ 0.*

We shall consider briefly what becomes of Theorem 17 when we consider the product $M^{p+q+r} = M^p \times S^q \times S^r$, $p \leqq q \leqq r$. Using the fundamental cycles of the three manifolds, define the three corresponding cycles, X^p, Y^q, Z^r as before. If $p < q < r$, we find by the above methods that $f(Z^{p+q+r}) \sim 0$ implies that one of $f(X^p)$, $f(Y^q)$, $f(Z^r)$ is ~ 0. Consider the case $p = q = r$. Say

---

[33] For a direct treatment, compare (24.18) below.

[34] For $K' = S^n$, the theorem was proved by H. Hopf, Fund. Math., vol. 25 (1935), pp. 427–440, part of Satz V. In the present form (with K' a manifold), it was communicated to me by Hopf in the letter referred to.

(24.15) $fX^p \sim \sum \xi_i X_i'^p, \quad fY^p \sim \sum \eta_i X_i'^p, \quad fZ^p \sim \sum \zeta_i X_i'^p.$

Let $C_1'^r, \cdots, C_s'^p$ form a base for p-cocycles in K'. Then if $fZ^{3p} \sim 0$, we have

(24.16) $$M_{ijk} = ((C_i'^p \smile C_j'^p) \smile C_k'^p) \cdot fZ^{3p} = 0 \qquad \text{(all } i, j, k\text{)}.$$

Define C^p, D^p, E^p as before. Then

(24.17) $$C^p \smile C^p \frown D^p \smile D^p \frown E^p \smile E^p \frown 0,$$
$$((C^p \smile D^p) \smile E^p) \cdot Z^{3p} = 1.$$

As $f'C_i'^p \frown \xi_i C^p + \eta_i D^p + \zeta_i E^p$ etc., the above relations give

(24.18) $$\begin{aligned} M_{ijk} &= \xi_i\eta_j\zeta_k(C^p \smile D^p \smile E^p)\cdot Z + \xi_i\zeta_j\eta_k(C^p \smile E^p \smile D^p)\cdot Z + \cdots \\ &= \xi_i(\eta_j\zeta_k + \epsilon\eta_k\zeta_j) + \xi_j(\eta_k\zeta_i + \epsilon\eta_i\zeta_k) + \xi_k(\eta_i\zeta_j + \epsilon\eta_j\zeta_i), \end{aligned}$$

where $\epsilon = (-1)^p$. Suppose p is even. Then by considering such quantities as M_{iii}, M_{iij}, we see that all the ξ_i or all the η_i or all the ζ_i vanish, so that one of fX^p, fY^p, fZ^p is ~ 0. If p is odd, then all determinants of the matrix of ξ_i, η_i and ζ_i vanish, so that fX^p, fY^p and fZ^p are linearly dependent under homology.

Appendix

Miscellaneous questions

25. The products in terms of other operations

We shall show how the products of §6 may be defined in terms of two other operations; one acts on single chains, and the other, on two chains of the same dimension.

Writing simplexes in their normal form §6, we define two homomorphisms of L^p into L^q, $q \leqq p$, by means of

(25.1) $$\xi_p^q(x_{i_0} \cdots x_{i_p}) = x_{i_0} \cdots x_{i_q}, \qquad \zeta_p^q(x_{i_0} \cdots x_{i_p})\, x_{i_{p-q}} \cdots x_{i_p}.$$

Let ξ_q^p and ζ_q^p $(q \leqq p)$ be the dual maps of L^q into L^p. Define also

(25.2) $$(\sum \alpha_i\sigma_i^p)\circ(\sum \beta_i\sigma_i^p) = \sum \alpha_i\beta_i\sigma_i^p.$$

Then (6.1) gives

(25.3) $$A^p \smile B^q = \xi_p^{p+q}A^p \circ \zeta_q^{p+q}B^q, \qquad A^q \frown B^{p+q} = \xi_{p+q}^p(\zeta_q^{p+q}A^q \circ B^{p+q}).$$

Let I^p be the sum of all (oriented) p-cells: $I^p = \sum x_{i_0} \cdots x_{i_p}$. Then clearly $I^0 = I$. Some relations following at once from the definitions are

(25.4) $$I^p\cdot(A\circ B) = A\cdot B, \qquad (A\circ B)\cdot C = A\cdot(B\circ C),$$

(25.5) $$\xi_q^p A^q\cdot B^p = A^q\cdot\xi_p^q B^p, \qquad (A \smile B)\cdot C = A\cdot(B \frown C).$$

(25.6) $$I^p\circ A^p = A^p\circ I^p = A^p, \qquad \xi_p^q I^p = \zeta_p^q I^p = I^q (q \geqq p),$$

(25.7) $$\delta I^{2p} = 0, \qquad \delta I^{2p+1} = I^{2p+2}.$$

For any subcomplex K' of K, let us interpret K' also as a chain of mixed dimension, namely, the sum of all its cells. Then $K = \sum I^p$, $K' \circ K' = K'$, and

(25.8) $\qquad A^p \circ K' = K' \circ A^p =$ the "part of A^p in K'".

A^p is in K' if and only if $A^p \circ K' = A^p$. Set $K'' = K - K'$. Then all the following twelve conditions are equivalent:

K' is closed, K'' is open, $\bar{K}' = K'$, $\mathrm{St}(K'') = K''$,

$$\sigma_i^p \circ K' = \sigma_i^p \quad \text{implies} \quad \partial\sigma_i^p \circ K' = \partial\sigma_i^p,$$

$$\sigma_i^p \circ K'' = \sigma_i^p \quad \text{implies} \quad \delta\sigma_i^p \circ K'' = \delta\sigma_i^p,$$

$$\partial(A \circ K') \circ K'' = 0, \qquad \delta(A \circ K'') \circ K' = 0,$$

$$\partial(A \circ K'') \circ K'' = \partial A \circ K'', \qquad \delta(A \circ K') \circ K' = \delta A \circ K',$$

$$\partial(A \circ K') \circ K' = \partial(A \circ K'), \qquad \delta(A \circ K'') \circ K'' = \delta(A \circ K'').$$

We may obtain certain chains related to vertices:

$$\zeta_0^p x_i \smile \xi_0^q x_i = \sum \text{all } (p+q)\text{-cells with } x_i \text{ as } (p+1)\text{th vertex},$$

$$\textstyle\sum_k \zeta_0^{p-k} x_i \smile \xi_0^k x_i = \sum \text{all } p\text{-cells with } x_i \text{ as a vertex}.$$

26. Resolution of chains into boundaries and cocycles

By considering the ranks of the incidence matrices, we may prove

THEOREM 18. *Using real or rational numbers as coefficients, any p-chain may be written uniquely as a p-boundary plus a p-cocycle, or as a p-coboundary plus a p-cycle.*

We shall prove the first statement. To prove uniqueness, suppose

$$\partial A_1 + B_1 = \partial A_2 + B_2 \qquad (\delta B_1 = \delta B_2 = 0).$$

Then setting $A = A_2 - A_1$, $B = B_1 - B_2 = \sum \beta_i \sigma_i^p$, (4.7) gives

$$B \cdot B = B \cdot \partial A = \delta B \cdot A = 0, \qquad \sum \beta_i^2 = 0, \qquad \text{each } \beta_i = 0.$$

Hence $B_1 = B_2$, and $\partial A_1 = \partial A_2$, as required. To prove the existence of the decomposition, consider the rank ρ^p and the numbers of rows and columns α^{p-1} and α^p of $\| {}^p\partial_j^i \|$; α^p is the number of p-cells in K. Clearly

(26.1) $\qquad \rho^{p+1} =$ number of independent p-boundaries,

(26.2) $\qquad \alpha^p - \rho^{p+1} =$ number of independent p-cocycles.

With real coefficients, L^p is a linear vector space of dimension α^p. By the above relations, the p-boundaries and p-cocycles form linear subspaces of dimensions ρ^{p+1} and $\alpha^p - \rho^{p+1}$. We saw that the subspaces were orthogonal; hence they generate L^p.

With integral coefficients, the p-boundaries and p-cocycles generate a subgroup M^p of L^p of rank α^p. Hence the difference group $L^p - M^p$ is finite,

and for some integer $m \neq 0$, $m\theta = 0$ for all θ in $L^p - M^p$. This means that *any chain mA^p is (uniquely) a boundary plus a cocycle.* From this the theorem with rational coefficients follows at once.

EXAMPLE. Let K consist of a, b, c, ab, bc, ca. Then using integer coefficients, a 0-chain A^0 is a boundary plus a cocycle if and only if $I \cdot A^0 \equiv 0 \pmod 3$, and a 1-chain A^1 is a coboundary plus a cycle if and only if $I^1 \cdot A^1 \equiv 0 \pmod 3$.

27. Dual maps and changes of base

We shall prove

THEOREM 19. *Let G and H be free groups, and let dual homomorphisms be defined in terms of the bases $x_1, \cdots, x_m$ in G and $y_1, \cdots, y_n$ in H. Then the bases $x_1', \cdots, x_m'$ in G and $y_1', \cdots, y_n'$ in H give the same definition of duality if and only if the new bases are formed from the original ones by permutations and changes of sign.*

Clearly any such changes of base are allowable. To prove the other half of the theorem, suppose

$$x_i' = \sum_j \alpha_{ij} x_j, \qquad x_i = \sum_j \alpha_{ij}^{-1} x_j', \qquad |\alpha_{ij}| = \pm 1,$$
$$y_i' = \sum_j \beta_{ij} y_j, \qquad y_i = \sum_j \beta_{ij}^{-1} y_j', \qquad |\beta_{ij}| = \pm 1.$$

Suppose both pairs of bases give the same definition of dual homomorphisms. Then if ϕ and ψ are dual, we may write

$$\phi x_i = \sum_j \phi_{ij} y_j, \qquad \psi y_i = \sum_j \phi_{ji} x_j,$$
$$\phi x_i' = \sum_j \phi_{ij}' y_j', \qquad \psi y_i' = \sum_j \phi_{ji}' x_j'.$$

These equations with the former ones give

$$\phi_{ij}' = \sum_{k,l} \alpha_{ik} \phi_{kl} \beta_{lj}^{-1}, \qquad \phi_{ji}' = \sum_{k,l} \beta_{ik} \phi_{lk} \alpha_{lj}^{-1}.$$

As these hold for all ϕ_{kl}, we may set $\phi_{kl} = 1$, and all other $\phi_{pq} = 0$. This gives

$$\alpha_{ik} \beta_{lj}^{-1} = \alpha_{ki}^{-1} \beta_{jl} \qquad \text{(all } i, j, k, l\text{)}. \tag{27.1}$$

Multiplying by $\alpha_{pk}\beta_{jq}$ and summing over j and k gives

$$\delta_{lq} \sum_k \alpha_{pk} \alpha_{ik} = \delta_{pi} \sum_j \beta_{jl} \beta_{jq} \qquad \text{(all } i, l, p, q\text{)}.$$

Giving p and q different values shows that

$$\sum_k \alpha_{pk} \alpha_{ik} = c\delta_{pi}, \qquad \sum_k \beta_{kp} \beta_{ki} = c\delta_{pi}, \tag{27.2}$$

for some number c. The left hand member is an element of the matrix product of $\| \alpha_{ij} \|$ by its transposed; hence its determinant is $|\alpha_{ij}|^2 = 1$. Therefore $c^m = 1$, and $c = \pm 1$; clearly $c = 1$. It follows that $\| \alpha_{ij} \|$ (and also $\| \beta_{ij} \|$) has a ± 1 in each column, the rest of the column being zeros. This proves the theorem.

Suppose we replace *free group* by *linear vector space*. Again we find (27.2); this time we know merely that $c > 0$. Hence *the bases may be altered by applying to each base an orthogonal transformation, and then multiplying each vector of each set by the same constant* $\neq 0$; *these are the only allowable alterations.*

BIBLIOGRAPHY

1. J. W. Alexander, *On the chains of a complex and their duals*, Proc. Nat. Ac. Sci., vol. 21 (1935), pp. 509–511.
2. J. W. Alexander, *On the ring of a compact metric space*, Proc. Nat. Ac. Sci., vol. 21 (1935), pp. 511–512.
3. J. W. Alexander, *On the connectivity ring of an abstract space*, Annals of Math., vol. 37 (1936), pp. 698–708.
4. P. Alexandroff-H. Hopf, *Topologie I*, Berlin, 1936.
5. E. Čech, *Multiplications on a complex*, Annals of Math., vol. 37 (1936), pp. 681–697.
6. H. Freudenthal, *Über Mannigfaltigkeiten und ihre Abbildungen*, Kon. Ak. v. Wet., Amsterdam, vol. 40 (1937), pp. 54–60. Also Annals of Math., vol. 38(1937), pp. 647–655 and 847–853. See also Comp. Math. vol. 5 (1937), pp. 315–318 (added in proof).
7. H. Hopf, *Zur Algebra der Abbildungen von Mannigfaltigkeiten*, Jour. f. Math., vol. 163 (1930), pp. 71–88.
8. A. Kolmogoroff, *Über die Dualität im Aufbau der kombinatorischen Topologie*, Recueil Math. Moscow, vol. 1 (43) (1936), pp. 97–102.
9. A. Kolmogoroff, *Homologierung des Komplexes und des local-bikompakten Raumes*, Recueil Math. Moscow, vol. 1 (43) (1936), pp. 701–705.
10. S. Lefschetz, *Topology*, Colloquium lectures, New York, 1930.
11. S. Lefschetz, *The rôle of algebra in topology*, Bull. Am. Math. Soc., vol. 43 (1937), pp. 345–359.
12. A. W. Tucker, *An abstract approach to manifolds*, Annals of Math., vol. 34 (1933), pp. 191–243.
13. H. Whitney, *On matrices of integers and combinatorial topology*, Duke Math. Jour., vol. 3 (1937), pp. 35–45.

Harvard University.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ON THE FOURIER COEFFICIENTS OF CERTAIN MODULAR FORMS OF POSITIVE DIMENSION

BY HANS RADEMACHER AND HERBERT S. ZUCKERMAN[1]

(Received October 13, 1937)

1. Introduction

One of us has recently obtained an exact formula for the number $p(n)$ of unrestricted partitions of n.[2] The proof, which makes use of the Hardy-Littlewood-Ramanujan method, rests essentially on the fact that the generating function

$$(1.1) \qquad f(x) = 1 + \sum_{n=1}^{\infty} p(n)x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)\cdots}, \qquad |x| < 1,$$

is very closely related to a certain modular form, usually called $\eta(\tau)$. We now propose to apply the new method to general modular forms of a type presently to be specified.

An analytic function $H(\omega_1, \omega_2)$ of two complex variables ω_1, ω_2 subject to the restriction $\Im(\omega_2/\omega_1) > 0$, is called a modular form if
(A) it is homogeneous:

$$H(\lambda\omega_1, \lambda\omega_2) = \lambda^r H(\omega_1, \omega_2),$$

the real number r being called its dimension;
(B) it is invariant with respect to all transformations

$$\omega_2' = a\omega_2 + b\omega_1$$
$$\omega_1' = c\omega_2 + d\omega_1$$

of the full modular group (or, a case which we do not consider in this paper, invariant under all transformations of a subgroup of the modular group):

$$H(\omega_1', \omega_2') = H(\omega_1, \omega_2).$$

Writing

$$\tau = \frac{\omega_2}{\omega_1}, \qquad \tau' = \frac{\omega_2'}{\omega_1'} = \frac{a\tau + b}{c\tau + d}$$

---

[1] National Research Fellow.

[2] *A convergent series for the partition function* $p(n)$, Proceedings of the National Academy of Sciences, vol. 23, no. 2., pp. 78–84 (1937); *On the partition function* $p(n)$, Proceedings of the London Mathematical Society, (2), vol. 43 (1937), pp. 241–154.

we get from (A) and (B)

$$H(1, \tau') = (c\tau + d)^{-r} H(1, \tau). \tag{1.21}$$

The function $H(1, \tau)$, in spite of the inhomogeneous notation, may also be called a modular form of dimension r. For non-integral values of r we should have to determine the branch of the many-valued function $(c\tau + d)^{-r}$. Instead of doing this we generalize the notion of modular forms by admitting a factor ϵ of absolute value 1 and independent of τ, such that we finally define a modular form of dimension r as a function $F(\tau)$ analytic in the upper τ-half-plane and satisfying a functional relation

$$F(\tau') = \epsilon \cdot (-i(c\tau + d))^{-r} F(\tau), \tag{1.22}$$

for every modular transformation

$$\tau' = \frac{a\tau + b}{c\tau + d} \tag{1.3}$$

of the modular group (or of one of its subgroups only). Since for $c \neq 0$ we can choose in (1.3) $c > 0$, we can assign

$$-\pi/2 < \arg(-i(c\tau + d)) < \pi/2 \tag{1.4}$$

and determine $(-i(c\tau + d))^{-r}$ as $|(c\tau + d)^{-r}| \exp(-ir \arg(-i(c\tau + d)))$. The multiplier ϵ, being independent of τ, depends only on the particular substitution (1.3) and may therefore, if necessary, also be written

$$\epsilon = \epsilon(a, b, c, d).$$

The case $c = 0$, which reduces (1.3) to

$$\tau' = \tau + b$$

is conveniently treated separately. We write

$$F(\tau + b) = \epsilon_b F(\tau), \tag{1.51}$$

and in particular

$$F(\tau + 1) = \epsilon_1 F(\tau) = e^{2\pi i\alpha} F(\tau) \tag{1.52}$$

where α can be restricted to

$$0 \leqq \alpha < 1. \tag{1.6}$$

From (1.52) we get

$$e^{-2\pi i\alpha(\tau+1)} F(\tau + 1) = e^{-2\pi i\alpha\tau} F(\tau),$$

from which we infer the validity of a Fourier expansion

$$e^{-2\pi i\alpha\tau} F(\tau) = \sum_{m=-\infty}^{+\infty} a_m e^{2\pi i m\tau}. \tag{1.7}$$

We now restrict our considerations to functions $F(\tau)$ whose developments (1.7) have at least one, but only a finite number of negative exponents m:

$$F(\tau) = e^{2\pi i a\tau} \sum_{m=-\mu}^{\infty} a_m e^{2\pi i m\tau}. \tag{1.71}$$

The function $e^{-2\pi i a\tau}F(\tau)$ has a pole at the parabolic point $\tau = i\infty$, if we measure the singularity by the uniformizing variable $x = e^{2\pi i\tau}$. The segment

$$P(e^{2\pi i\tau}) = \sum_{m=-\mu}^{-1} a_m e^{2\pi i m\tau} \tag{1.8}$$

is the "principal part" of the function at that singularity. We assume moreover $F(\tau)$ to be regular at all finite points of the upper half-plane, so that the series (1.71) is convergent for $\Im(\tau) > 0$.

Our method requires eventually for its applicability the positivity of the dimension r (cf. end of no. 4).

It is now our aim to determine the coefficients a_m for $m \geqq 0$ as they arise from the combination of two sources, viz. the functional relation (1.22), (1.52) on one side, and the principal part (1.8) of $e^{-2\pi i a\tau}F(\tau)$ at the parabolic point $\tau = i\infty$ of the fundamental region on the other side; P itself is given by the coefficients $a_{-\mu}, \cdots, a_{-1}$.

It is convenient to introduce the variable $x = e^{2\pi i\tau}$ and to put

$$e^{-2\pi i a\tau}F(\tau) = f(x) = \sum_{m=-\mu}^{\infty} a_m x^m. \tag{1.9}$$

The function $f(x)$ is analytic inside the unit circle and has a pole of order μ at $x = 0$. The rational function

$$P(x) = a_{-\mu}x^{-\mu} + \cdots + a_{-1}x^{-1} \tag{1.91}$$

is the principal part of $f(x)$ at that pole.[3]

---

[3] It may be noticed that our present notation is slightly at variance with the notation used in (1.1) and that adopted in the papers cited in the preceding footnote. There we had as $F(\tau)$ the special function $\eta(\tau)^{-1}$. Writing it in the form (1.71) we have

$$\eta(\tau)^{-1} = e^{23\pi i\tau/12}\{e^{-2\pi i\tau} + a_0 + a_1 e^{2\pi i\tau} + \cdots\}$$

and hence in analogy to (1.9)

$$\begin{aligned} e^{-23\pi i\tau/12}\eta(\tau)^{-1} = f(x) &= x^{-1} + a_0 + a_1 x + \cdots \\ &= x^{-1}(1 + a_0 x + a_1 x^2 + \cdots) \\ &= x^{-1}\left(1 + \sum_{n=1}^{\infty} p(n)x^n\right). \end{aligned}$$

Our present function $f(x)$ therefore differs by a factor x^{-1} from that which is called $f(x)$ in (1.1) and the other places mentioned above.

2. Cauchy integral and Farey dissection

We have now, by Cauchy's integral formula,

$$a_m = \frac{1}{2\pi i}\int_C \frac{f(x)}{x^{m+1}}\,dx,$$

where C may be the circle

$$|x| = e^{-2\pi N^{-2}}$$

and N a positive integer. We make the usual Farey dissection of the circle C into arcs $\xi_{h,k}$, using the Farey series of order N. Thus we have

$$a_m = \sum_{\substack{h,k \\ 0\le h<k\le N \\ (h,k)=1}} \frac{1}{2\pi i}\int_{\xi_{h,k}} \frac{f(x)}{x^{m+1}}\,dx. \tag{2.1}$$

We can describe the Farey arc $\xi_{h,k}$ by

$$x = \exp\left(-2\pi N^{-2} + 2\pi i h/k + 2\pi i\varphi\right) \tag{2.21}$$

$$-\vartheta'_{h,k} \le \varphi \le \vartheta''_{h,k} \tag{2.22}$$

and get through the substitution (2.21)

$$a_m = e^{2\pi N^{-2}m} \sum_{\substack{h,k \\ 0\le h<k\le N \\ (h,k)=1}} e^{-2\pi i h m/k} \times \int_{-\vartheta'}^{\vartheta''} f\left[\exp\left(2\pi i h/k - 2\pi(N^{-2} - i\varphi)\right)\right] e^{-2\pi i m\varphi}\,d\varphi, \tag{2.3}$$

where we have briefly written ϑ', ϑ'' instead of $\vartheta'_{h,k}$, $\vartheta''_{h,k}$.

In order to make use of the transformation formula (1.22) we write, for $c > 0$,

$$-i(c\tau + d) = z, \qquad \tau = \frac{iz}{c} - \frac{d}{c}, \qquad \Re(z) > 0,$$

and choose the modular transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} h' & -\dfrac{hh'+1}{k} \\ k & -h \end{pmatrix}$$

where h' is a solution of $hh' \equiv -1 \pmod{k}$. We then have

$$\tau = \frac{iz}{k} + \frac{h}{k}, \qquad \tau' = \frac{i}{kz} + \frac{h'}{k}$$

and therefore, by (1.22),

$$F\left(\frac{i}{kz} + \frac{h'}{k}\right) = \epsilon z^{-r} F\left(\frac{iz}{k} + \frac{h}{k}\right)$$

and from (1.9)

$$f(e^{-2\pi(z-ih)/k}) = \epsilon^{-1} z^r e^{2\pi a(z-ih)/k} e^{2\pi i a h'/k} e^{-2\pi a/kz} f(e^{2\pi ih'/k} e^{-2\pi/kz}), \tag{2.4}$$

where

$$\epsilon = \epsilon(h', -(hh'+1)/k, k, -h).$$

We could make h' unique by adding the restriction $0 \leqq h' < k$, but this is unessential, since (2.4) shows that the expression

$$\Omega_{h,k} = \epsilon^{-1} e^{2\pi i a(h'-h)/k} \tag{2.5}$$

must be independent of the special choice of h'.[4]

Introducing (2.5) and

$$\Psi_k(z) = z^r e^{2\pi a(z-1/z)/k} \tag{2.6}$$

we can replace (2.4) by

$$f(e^{-2\pi(z-ih)/k}) = \Omega_{h,k} \Psi_k(z) f(e^{2\pi ih'/k} e^{-2\pi/kz}).$$

This equation applied to (2.3) yields

$$a_m = e^{2\pi N^{-2} m} \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi ihm/k}$$
$$\int_{-\vartheta'}^{\vartheta''} \Psi_k(k(N^{-2} - i\varphi)) f(\exp[(2\pi/k)(ih' - k^{-1}(N^{-2} - i\varphi)^{-1})]) e^{-2\pi im\varphi} d\varphi. \tag{2.7}$$

3. An estimation

We are now prepared to make full use of the fact that $f(x)$ in the neighborhood of $x = 0$ is dominated by the principal part $P(x)$. For that purpose we split (2.7) into two parts

$$a_m = Q(m) + R(m) \tag{3.1}$$

where

$$Q(m) = e^{2\pi N^{-2} m} \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi ihm/k}$$
$$\int_{-\vartheta'}^{\vartheta''} \Psi_k(k(N^{-2} - i\varphi)) P(\exp[(2\pi/k)(ih' - k^{-1}(N^{-2} - i\varphi)^{-1})]) e^{-2\pi im\varphi} d\varphi \tag{3.2}$$

---

[4] This means that the ϵ's must, for consistency of (1.22) and (1.52), satisfy (among others) the relation

$$\epsilon(a, b, c, d) = \epsilon(a + c, b + d, c, d) e^{-2\pi i a}, \quad c > 0.$$

and

$$R(m) = e^{2\pi N^{-2}m} \sum_{\substack{h,k \\ 0\leqq h<k\leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i h m/k}$$

(3.3)

$$\int_{-\vartheta'}^{\vartheta''} \Psi_k(k(N^{-2} - i\varphi)) D\,(\exp[(2\pi/k)(ih' - k^{-1}(N^{-2} - i\varphi)^{-1})]) e^{-2\pi i m\varphi} d\varphi$$

with

$$D(x) = f(x) - P(x) = \sum_{m=0}^{\infty} a_m x^m. \tag{3.4}$$

We are now going to make an estimate of $R(m)$. From the theory of Farey fractions we have

$$1/2kN \leqq \vartheta' \leqq 1/kN, \qquad 1/2kN \leqq \vartheta'' \leqq 1/kN$$

and therefore, since $k \leqq N$, we find for $-\vartheta' \leqq \varphi \leqq \vartheta''$

$$\Re(k(N^{-2} - i\varphi)) = kN^{-2},$$

$$\Re\left(\frac{1}{k(N^{-2} - i\varphi)}\right) = \frac{N^{-2}}{k(N^{-4} + \varphi^2)} \geqq \frac{N^{-2}}{k(N^{-4} + k^{-2}N^{-2})} = \frac{k}{k^2N^{-2} + 1} \geqq \frac{k}{2},$$

$$|\,k(N^{-2} - i\varphi)\,| = k(N^{-4} + \varphi^2)^{\frac{1}{2}} \leqq (k^2N^{-4} + N^{-2})^{\frac{1}{2}} \leqq 2^{\frac{1}{2}}N^{-1}.$$

Using these results we have by (2.6), (3.4)

$$\begin{aligned} &|\,\Psi_k(k(N^{-2} - i\varphi)) \cdot D\,(\exp[(2\pi/k)(ih' - k^{-1}(N^{-2} - i\varphi)^{-1})])\,| \\ &\qquad \leqq 2^{r/2}N^{-r} \exp[2\pi\alpha(N^{-2} - k^{-1}\Re(k^{-1}(N^{-2} - i\varphi)^{-1}))] \\ &\qquad \times \sum_{m=0}^{\infty} |\,a_m\,| \exp[-2\pi m\Re(k^{-2}(N^{-2} - i\varphi)^{-1})] \\ &\qquad \leqq 2^{r/2}N^{-r}e^{2\pi\alpha N^{-2}}e^{-\pi\alpha} \sum_{m=0}^{\infty} |\,a_m\,| e^{-\pi m} = CN^{-r}e^{2\pi\alpha N^{-2}}, \end{aligned} \tag{3.5}$$

where

$$C = 2^{r/2}e^{-\pi\alpha} \sum_{m=0}^{\infty} |\,a_m\,| e^{-\pi m}$$

is finite since $|\,e^{-\pi}\,| < 1$ and the series (1.9) is convergent inside the unit circle. Combining (3.5) with (3.3) we have

$$|\,R(m)\,| \leqq e^{2\pi N^{-2}m} \sum_{\substack{h,k \\ 0\leqq h<k\leqq N \\ (h,k)=1}} CN^{-r}e^{2\pi\alpha N^{-2}} \int_{-\vartheta'}^{\vartheta''} d\varphi = CN^{-r}e^{2\pi(m+\alpha)N^{-2}}. \tag{3.6}$$

4. A convergent series for a_m

We now evaluate $Q(m)$, under the condition

$$m + \alpha > 0, \tag{4.1}$$

which will be necessary for the convergence of a certain integral. The condition (4.1) admits, because of (1.6), all $m \geqq 0$, with the single exception $m = 0$ for $\alpha = 0$, a case which we shall have to treat separately.

Making the substitution

$$w = N^{-2} - i\varphi$$

in (3.2) we have

$$Q(m) = \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i h m/k} \frac{1}{i} \int_{N^{-2}-i\vartheta''}^{N^{-2}+i\vartheta'} \Psi_k(kw) P(e^{2\pi i h'/k} e^{-2\pi/k^2 w}) e^{2\pi m w}\, dw$$

which, by (2.6) and (1.91), can be written as

$$Q(m) = \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i h m/k} k^r \sum_{\nu=1}^{\mu} a_{-\nu} e^{-2\pi i h'\nu/k} I_k(m, \nu), \tag{4.21}$$

where

$$I_k(m, \nu) = \frac{1}{i} \int_{N^{-2}-i\vartheta''}^{N^{-2}+i\vartheta'} w^r \exp\,[2\pi((m + \alpha)w + (\nu - \alpha)k^{-2}w^{-1})]\, dw, \tag{4.22}$$

$$\nu = 1, \cdots, \mu.$$

We cut the complex w-plane from 0 to $-\infty$ along the negative real axis and consider a path of integration encircling the cut in the positive sense and connecting the points

$$-\infty, -\epsilon, -\epsilon - i\vartheta'', N^{-2} - i\vartheta'', N^{-2} + i\vartheta', -\epsilon + i\vartheta', -\epsilon, -\infty \tag{4.3}$$

by straight lines. Then we can write

$$\begin{aligned} I_k(m, \nu) = {} & \frac{1}{i} \int_{-\infty}^{(0+)} - \frac{1}{i} \int_{-\infty}^{-\epsilon} - \frac{1}{i} \int_{-\epsilon}^{-\epsilon-i\vartheta''} - \frac{1}{i} \int_{-\epsilon-i\vartheta''}^{N^{-2}-i\vartheta''} - \frac{1}{i} \int_{N^{-2}-i\vartheta''}^{-\epsilon+i\vartheta'} \\ & - \frac{1}{i} \int_{-\epsilon+i\vartheta'}^{-\epsilon} - \frac{1}{i} \int_{-\epsilon}^{-\infty} = L_k(m, \nu) - J_1 - J_2 - J_3 - J_4 - J_5 - J_6, \end{aligned} \tag{4.41}$$

say. All these integrals have the same integrand

$$w^r \exp\,[2\pi((m + \alpha)w + (\nu - \alpha)k^{-2}w^{-1})]. \tag{4.42}$$

The integral J_1 is to be taken on the border below the cut, J_6 above the cut. We assume moreover

$$0 < \epsilon < N^{-2}.$$

In the integral J_2 we have

$$w = -\epsilon + iv, \qquad 0 \geqq v \geqq -\vartheta'',$$

$$\Re(w) = -\epsilon, \qquad \Re\left(\frac{1}{w}\right) = \frac{-\epsilon}{\epsilon^2 + v^2} < 0,$$

$$|w| = (\epsilon^2 + v^2)^{\frac{1}{2}} \leqq (N^{-4} + k^{-2}N^{-2})^{\frac{1}{2}} \leqq 2^{\frac{1}{2}}k^{-1}N^{-1},$$

and therefore

$$|J_2| \leqq \vartheta'' 2^{r/2}k^{-r}N^{-r}e^{-2\pi(m+\alpha)\epsilon} < 2^{r/2}k^{-r-1}N^{-r-1}. \tag{4.51}$$

Similarly we have

$$|J_5| < 2^{r/2}k^{-r-1}N^{-r-1}. \tag{4.52}$$

In the integral J_3 we have

$$w = u - i\vartheta'', \qquad -N^{-2} < -\epsilon \leqq u \leqq N^{-2},$$

$$\Re(w) = u \leqq N^{-2}, \qquad \Re\left(\frac{1}{w}\right) = \frac{u}{u^2 + \vartheta''^2} \leqq \frac{N^{-2}}{\vartheta''^2} \leqq 4k^2,$$

$$|w| = (u^2 + \vartheta''^2)^{\frac{1}{2}} \leqq (N^{-4} + N^{-2}k^{-2})^{\frac{1}{2}} \leqq 2^{\frac{1}{2}}k^{-1}N^{-1}$$

and therefore

$$\begin{aligned} |J_3| &\leqq (N^{-2} + \epsilon)2^{r/2}k^{-r}N^{-r}\exp\left[2\pi(m+\alpha)N^{-2} + 8\pi(\nu - \alpha)\right] \\ &\leqq 2^{1+r/2}k^{-r-1}N^{-r-1}\exp\left[2\pi(m+\alpha)N^{-2} + 8\pi(\nu - \alpha)\right], \end{aligned} \tag{4.61}$$

and similarly

$$|J_4| \leqq 2^{1+r/2}k^{-r-1}N^{-r-1}\exp\left[2\pi(m+\alpha)N^{-2} + 8\pi(\nu - \alpha)\right]. \tag{4.62}$$

Finally we have

$$\begin{aligned} J_1 + J_6 &= \frac{1}{i}\int_{-\infty}^{-\epsilon} |w|^r e^{-\pi i r}\exp\left[2\pi(m+\alpha)w + 2\pi(\nu - \alpha)k^{-2}w^{-1}\right]dw \\ &\quad + \frac{1}{i}\int_{-\epsilon}^{-\infty} |w|^r e^{\pi i r}\exp\left[2\pi(m+\alpha)w + 2\pi(\nu - \alpha)k^{-2}w^{-1}\right]dw \\ &= -2\sin\pi r\int_{\epsilon}^{\infty} t^r\exp\left[-2\pi(m+\alpha)t - 2\pi(\nu - \alpha)k^{-2}t^{-1}\right]dt, \end{aligned} \tag{4.7}$$

where the condition (4.1) ensures the convergence of the integrals. Combining (4.41), (4.51), (4.52), (4.61), (4.62), (4.7) and making $\epsilon \to 0$, we have

$$\begin{aligned} I_k(m,\nu) &= L_k(m,\nu) + 2\sin\pi r\int_0^{\infty} t^r\exp\left[-2\pi(m+\alpha)t - 2\pi(\nu - \alpha)k^{-2}t^{-1}\right]dt \\ &\quad + 6\Theta\, 2^{r/2}e^{8\pi(\mu-\alpha)}k^{-r-1}N^{-r-1}e^{2\pi(m+\alpha)N^{-2}}, \qquad |\Theta| < 1, \qquad \nu = 1, \cdots, \mu. \end{aligned}$$

Introducing this into (4.21) we obtain

$$Q(m) = \sum_{\substack{h,k \\ 0\leqq h<k\leqq N \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i h m/k} k^r \sum_{\nu=1}^{\mu} a_{-\nu} e^{-2\pi i h'\nu/k} \{L_k(m,\nu) + H_k(m,\nu)\}$$

(4.81)

$$+ 6\Theta\, 2^{r/2} e^{8\pi(\mu-\alpha)} N^{-r-1} e^{2\pi(m+\alpha)N^{-2}} \sum_{\nu=1}^{\mu} |a_{-\nu}| \sum_{\substack{h,k \\ 0\leqq h<k\leqq N \\ (h,k)=1}} k^{-1}, \qquad |\Theta| < 1,$$

with

$$L_k(m,\nu) = \frac{1}{i}\int_{-\infty}^{(0+)} w^r \exp\left[2\pi(m+\alpha)w + 2\pi(\nu-\alpha)k^{-2}w^{-1}\right] dw, \tag{4.82}$$

$$H_k(m,\nu) = 2\sin\pi r \int_0^{\infty} t^r \exp\left[-2\pi(m+\alpha)t - 2\pi(\nu-\alpha)k^{-2}t^{-1}\right] dt. \tag{4.83}$$

Now we have in (4.81)

$$\sum_{\substack{h,k \\ 0\leqq h<k\leqq N \\ (h,k)=1}} k^{-1} \leqq \sum_{\substack{k \\ 0<k\leqq N}} 1 = N,$$

and thus (4.81), (3.1) and (3.6) together give the result

$$a_m = \sum_{k=1}^{N} k^r \sum_{\substack{0\leqq h<k \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i h m/k} \sum_{\nu=1}^{\mu} a_{-\nu} e^{-2\pi i h'\nu/k} \{L_k(m,\nu) + H_k(m,\nu)\}$$

$$+ O(N^{-r} e^{2\pi(m+\alpha)N^{-2}}).$$

If we keep m fixed and let N tend to infinity, the error term tends to zero, since $r > 0$. It is at this point that we must definitely make use of the positivity of the dimension r of our modular form $F(\tau)$.

The limit 0 of the error term signifies the convergence of the series thus obtained for a_m :

$$a_m = \sum_{k=1}^{\infty} k^r \sum_{\nu=1}^{\mu} a_{-\nu} A_{k,\nu}(m) \{L_k(m,\nu) + H_k(m,\nu)\} \tag{4.91}$$

with

$$A_{k,\nu}(m) = \sum_{\substack{0\leqq h<k \\ (h,k)=1}} \Omega_{h,k} e^{-2\pi i(hm+h'\nu)/k}. \tag{4.92}$$

5. Evaluation of the integrals

The integrals (4.82) and (4.83) are well-known in the theory of Bessel functions. Indeed we have[5]

$$I_\rho(z) = \frac{(z/2)^\rho}{2\pi i}\int_{-\infty}^{(0+)} t^{-\rho-1}\exp\left(t + z^2/4t\right) dt \tag{5.11}$$

[5] G. N. Watson, *Theory of Bessel Functions* (Cambridge 1922), p. 181, (1), p. 183, (15).

and

$$K_\rho(z) = \frac{(z/2)^\rho}{2} \int_0^\infty t^{-\rho-1} \exp(-t - z^2/4t)\, dt, \tag{5.12}$$

which we need here only for positive z and real ρ. The functions $I_\rho(z)$ and $K_\rho(z)$ are the Bessel functions of the first kind and the third kind respectively, with purely imaginary arguments. Comparing (4.82) and (4.83) with (5.11) and (5.12) we get

$$L_k(m, \nu) = \frac{2\pi}{k^{r+1}} \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{-r-1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}})$$

$$H_k(m, \nu) = \frac{4 \sin \pi r}{k^{r+1}} \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} K_{-r-1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}).$$

But[6]

$$\sin\pi\rho\, K_\rho(z) = (\pi/2)(I_{-\rho}(z) - I_\rho(z)),$$

and so we have, since $\sin\pi r = \sin(-\pi(r + 1))$,

$$H_k(m, \nu) = \frac{2\pi}{k^{r+1}} \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} \{I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}) - I_{-r-1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}})\},$$

and hence

$$L_k(m, \nu) + H_k(m, \nu) = \frac{2\pi}{k^{r+1}} \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}).$$

This reduces (4.91) to

$$a_m = 2\pi \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\nu=1}^{\mu} a_{-\nu} A_{k,\nu}(m) \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}). \tag{5.2}$$

Each term of this infinite series consists of μ members. The summation signs in (5.2) can, however, be interchanged, so that we get the sum of μ convergent series. Indeed, a single series

$$\sum_{k=1}^{\infty} \frac{1}{k} A_{k,\nu}(m) \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}) \tag{5.3}$$

is absolutely convergent, since we have on one hand, from (4.92),

$$\left|\frac{1}{k} A_{k,\nu}(m)\right| \leqq 1,$$

[6] Watson, l. c., p. 78, (6).

and on the other hand[7]

$$\left(\frac{\nu-\alpha}{m+\alpha}\right)^{(r+1)/2} I_{r+1}\left((4\pi/k)(\nu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}\right) = \frac{(2\pi(\nu-\alpha))^{r+1}}{k^{r+1}} \sum_{l=0}^{\infty} \frac{(2\pi/k)^{2l}(\nu-\alpha)^l(m+\alpha)^l}{l!\,\Gamma(r+l+2)}. \tag{5.4}$$

The series (5.3) converges therefore absolutely for $r > 0$. It follows that (5.2) can be changed into

$$a_m = 2\pi \sum_{\nu=1}^{\mu} a_{-\nu} \sum_{k=1}^{\infty} \frac{1}{k} A_{k,\nu}(m) \left(\frac{\nu-\alpha}{m+\alpha}\right)^{(r+1)/2} I_{r+1}\left((4\pi/k)(\nu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}\right). \tag{5.5}$$

We finally notice that because of (2.5) and (4.92) the coefficients $A_{k,\nu}(m)$ can be written as

$$A_{k,\nu}(m) = \sum_{\substack{0\leq h<k \\ (h,k)=1}} \epsilon\left(h', -\frac{hh'+1}{k}, k, -h\right)^{-1} \exp\left[(-2\pi i/k)((\nu-\alpha)h' + (m+\alpha)h)\right]. \tag{5.6}$$

6. Connection between r and α

Up to the present we had assumed that $m + \alpha > 0$. We shall show, however, that (5.5) remains true for

$$m + \alpha = 0, \tag{6.11}$$

which means

$$\alpha = 0, \qquad m = 0, \tag{6.12}$$

if we interpret (5.4) in the proper way as

$$\lim_{m+\alpha\to 0} \left(\frac{\nu-\alpha}{m+\alpha}\right)^{(r+1)/2} I_{r+1}\left((4\pi/k)(\nu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}\right) = \frac{(2\pi\nu)^{r+1}}{k^{r+1}\Gamma(r+2)}. \tag{6.2}$$

For this purpose we have to study the connection between r and α.

For the two generating transformations

$$S\colon \tau' = \tau + 1,$$

$$T\colon \tau' = -\frac{1}{\tau}$$

we have respectively

$$F(\tau+1) = e^{2\pi i\alpha} F(\tau), \tag{6.31}$$

$$F\left(-\frac{1}{\tau}\right) = \epsilon_0(-i\tau)^{-r} F(\tau) \tag{6.32}$$

[7] Watson, l.c., p. 77, (2).

with

$$\epsilon_0 = \epsilon(0, -1, 1, 0).$$

Replacing in (6.32) τ by $-1/\tau$ we get

$$F(\tau) = \epsilon_0 \left(\frac{i}{\tau}\right)^{-r} F\left(-\frac{1}{\tau}\right)$$

$$= \epsilon_0^2 \left(\frac{i}{\tau}\right)^{-r} (-i\tau)^{-r} F(\tau) \tag{6.41}$$

$$= \epsilon_0^2 \left|\frac{1}{\tau}\right|^{-r} |\tau|^{-r} \exp\left\{-ir\left(\arg\frac{i}{\tau} + \arg(-i\tau)\right)\right\} F(\tau).$$

From (1.4) we infer

$$-\pi/2 < \arg\frac{i}{\tau} < \pi/2$$

$$-\pi/2 < \arg(-i\tau) < \pi/2$$

and hence

$$-\pi < \arg\frac{i}{\tau} + \arg(-i\tau) < \pi. \tag{6.42}$$

But we have

$$\arg\frac{i}{\tau} + \arg(-i\tau) \equiv \arg\left(\frac{i}{\tau}\cdot(-i\tau)\right) \equiv 0 \pmod{2\pi},$$

which together with (6.42) leaves only the possibility

$$\arg\frac{i}{\tau} + \arg(-i\tau) = 0.$$

Thus (6.41) goes over into

$$F(\tau) = \epsilon_0^2 F(\tau),$$

and we get

$$\epsilon_0^2 = 1. \tag{6.5}$$

From (6.31) and (6.32) we deduce

$$F\left(1 - \frac{1}{\tau}\right) = e^{2\pi i\alpha}\epsilon_0(-i\tau)^{-r}F(\tau).$$

If we replace, in this equation, two times in succession τ by $1 - 1/\tau$ we obtain

$$F(\tau) = e^{6\pi i\alpha}\epsilon_0^3\left(\frac{i}{\tau-1}\right)^{-r}\left(-i\left(1-\frac{1}{\tau}\right)\right)^{-r}(-i\tau)^{-r}F(\tau)$$

$$= e^{6\pi i\alpha}\epsilon_0\left|\frac{1}{\tau-1}\right|^{-r}\left|1-\frac{1}{\tau}\right|^{-r}|\tau|^{-r} \tag{6.61}$$

$$\times \exp\left\{-ir\left(\arg\frac{i}{\tau-1} + \arg\left(-i\left(1-\frac{1}{\tau}\right)\right) + \arg(-i\tau)\right)\right\}F(\tau),$$

where we have made use of (6.5). Application of (1.4) gives

$$(6.62)\qquad -\frac{3\pi}{2} < \arg\frac{i}{\tau-1} + \arg\left(-i\left(1-\frac{1}{\tau}\right)\right) + \arg(-i\tau) < \frac{3\pi}{2}.$$

On the other hand we have

$$\arg\frac{i}{\tau-1} + \arg\left(-i\left(1-\frac{1}{\tau}\right)\right) + \arg(-i\tau)$$
$$\equiv \arg\left(\frac{i}{\tau-1}\cdot\frac{-i(\tau-1)}{\tau}\cdot(-i\tau)\right) \equiv \arg(-i) \equiv -\frac{\pi}{2} \pmod{2\pi}.$$

From this and (6.62) we get

$$\arg\frac{i}{\tau-1} + \arg\left(-i\left(1-\frac{1}{\tau}\right)\right) + \arg(-i\tau) = -\frac{\pi}{2},$$

and therefore (6.61) becomes

$$(6.7)\qquad F(\tau) = e^{6\pi i\alpha}\epsilon_0 e^{\pi i r/2}F(\tau),$$
$$\epsilon_0 = e^{-6\pi i\alpha - \pi i r/2}.$$

Since, by (6.5), $\epsilon_0 = \pm 1$, we conclude from (6.7) that

$$6\alpha + r/2$$

is always an integer.

This result shows in particular that $\alpha = 0$ in (6.12) is necessarily connected with an integer r (which is moreover even).

7. The exceptional case $m = \alpha = 0$; the main theorem

Now if r has an integral value the reasoning of no. 4 may be considerably simplified. In this case the integrand (4.42)

$$w^r \exp\left[2\pi(m+\alpha)w + 2\pi(\nu-\alpha)k^{-2}w^{-1}\right]$$

is unique in the whole w-plane, and we need not cut the plane along the negative real axis. Instead of using the loop path indicated by (4.3) we can integrate around $w = 0$ along a rectangle R with the vertices

$$-\epsilon - i\vartheta'', \qquad N^{-2} - i\vartheta'', \qquad N^{-2} + i\vartheta', \qquad -\epsilon + i\vartheta'$$

and get

$$I_k(m,\nu) = \frac{1}{i}\int_R - \frac{1}{i}\int_{N^{-2}+i\vartheta'}^{-\epsilon+i\vartheta'} - \frac{1}{i}\int_{-\epsilon+i\vartheta'}^{-\epsilon-i\vartheta''} - \frac{1}{i}\int_{-\epsilon-i\vartheta''}^{N^{-2}-i\vartheta''}$$
$$= L_k^*(m,\nu) - J_1^* - J_2^* - J_3^*,$$

say. Here J_1^* corresponds to J_4, J_3^* to J_3, and J_2^* to $J_2 + J_5$ in (4.41). The estimations are completely analogous to those in no. 4, only that we can now

disregard the convergence condition $m + \alpha > 0$, since no improper integrals appear. In this way we get instead of (4.91)

$$a_m = \sum_{k=1}^{\infty} k^r \sum_{\nu=1}^{\mu} a_{-\nu} A_{k,\nu}(m) L_k^*(m, \nu).$$

We are interested only in the case $m = 0$, $\alpha = 0$:

$$a_0 = \sum_{k=1}^{\infty} k^r \sum_{\nu=1}^{\mu} a_{-\nu} A_{k,\nu}(0) L_k^*(0, \nu)$$

with

$$L_k^*(0, \nu) = \frac{1}{i} \int^{(0+)} w^r e^{2\pi\nu/k^2 w}\, dw = \frac{2\pi}{(r+1)!} \left(\frac{2\pi\nu}{k^2}\right)^{r+1}.$$

This yields, after interchanging the summations with respect to k and ν

$$a_0 = 2\pi \sum_{\nu=1}^{\mu} a_{-\nu} \sum_{k=1}^{\infty} \frac{1}{k} A_{k,\nu}(0) \frac{1}{(r+1)!} \left(\frac{2\pi\nu}{k}\right)^{r+1}, \tag{7.1}$$

which is exactly the formula which was to be expected in view of (5.5) and (6.2).

We collect our results in the following

THEOREM 1: *Let $F(\tau)$ be a modular form of positive dimension r, regular in the upper τ-half-plane, and satisfying*

$$F\left(\frac{a\tau + b}{c\tau + d}\right) = \epsilon(a, b, c, d) \cdot (-i(c\tau + d))^{-r} F(\tau), \qquad c > 0,$$

$$F(\tau + 1) = e^{2\pi i \alpha} F(\tau), \qquad 0 \leqq \alpha < 1.$$

Also let its Fourier expansion contain only a finite number of terms with negative exponents:

$$F(\tau) = e^{2\pi i \alpha \tau} \sum_{m=-\mu}^{\infty} a_m e^{2\pi i m \tau}.$$

Then the coefficients a_m for $m \geqq 0$ are determined by those with $-\mu \leqq m \leqq -1$ through the formula

$$a_m = 2\pi \sum_{\nu=1}^{\infty} a_{-\nu} \sum_{k=1}^{\mu} \frac{1}{k} A_{k,\nu}(m) \left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}), \tag{7.2}$$

where $A_{k,\nu}(m)$ is given by (5.6). *For $\alpha = m = 0$ this formula is to be understood as meaning* (7.1).

We see in (7.2) that each a_m for $m \geqq 0$ depends linearly on $a_{-1}, \cdots a_{-\mu}$. Our analysis would hold true also for $a_{-1} = \cdots = a_{-\mu} = 0$, in which case consequently all a_m for $m \geqq 0$ would vanish. But $a_{-1} = \cdots = a_{-\mu} = 0$ means, as the Fourier expansion shows, that $F(\tau)$ tends to a finite value (different from 0

or not) when $\Im(\tau) \to \infty$, in which case we call $F(\tau)$ *regular at infinity*. We have therefore the

COROLLARY: *A modular form, belonging to the full modular group, of positive dimension, and regular at all finite points of the upper half-plane as well as at infinity, must necessarily vanish identically.*

8. Parametric representation of the modular forms under discussion

We can completely characterize the class of all functions $F(\tau)$ which have the properties (1.22), (1.52), and a principal part (1.8) at $\tau = i\infty$.

The function

$$\eta(\tau) = e^{\pi i\tau/12} \prod_{n=1}^{\infty} (1 - e^{2\pi i n\tau}) \tag{8.11}$$

is different from zero for all τ in the upper half-plane. Therefore for any real r the function

$$\eta(\tau)^{2r} = e^{2\pi i r\tau/12} \prod_{n=1}^{\infty} (1 - e^{2\pi i n\tau})^{2r} \tag{8.12}$$

can be uniquely defined if we take for each factor that branch which tends to 1 with $\tau \to i\infty$. We have then

$$\eta(\tau + 1)^{2r} = e^{2\pi i r/12}\eta(\tau)^{2r} \tag{8.21}$$

and

$$\eta\left(-\frac{1}{\tau}\right)^{2r} = (-i\tau)^r \eta(\tau)^{2r}, \tag{8.22}$$

where (8.21) is an immediate consequence of (8.12), and where (8.22) follows from the transformation equation for $\log \eta(\tau)$, first given by Dedekind.[8] The function

$$G(\tau) = F(\tau) \cdot \eta(\tau)^{2r} \tag{8.3}$$

has, as consequence of (6.31), (6.32), (8.21), (8.22) the properties

$$G(\tau + 1) = e^{2\pi i(\alpha + r/12)} G(\tau),$$

$$G\left(-\frac{1}{\tau}\right) = \epsilon_0 G(\tau).$$

But from (6.7) we know that

$$q = 6\alpha + r/2 \tag{8.41}$$

[8] *Erläuterungen zu den Fragmenten über die Grenzfälle der elliptischen Modulfunktionen*, Riemanns Werke, 2. Aufl., pp. 466–478 especially formula (17).

is always an integer, and we can write

$$G(\tau + 1) = e^{\pi i q/3} G(\tau), \tag{8.42}$$

$$G\left(-\frac{1}{\tau}\right) = e^{\pi i q} G(\tau). \tag{8.43}$$

The function $G(\tau)$ is therefore a modular form of dimension zero, i.e. a "modular function." A consequence of the last two equations is

$$G\left(1 - \frac{1}{\tau}\right) = e^{4\pi i q/3} G(\tau). \tag{8.44}$$

By expanding around the point $\tau = i$ we can see from (8.43) that, for odd q, $G(\tau)$ must have a zero of odd order. For this purpose we put

$$t = \frac{\tau - i}{\tau + i}, \qquad \tau = -i\frac{t+1}{t-1},$$

$$G(\tau) = G\left(-i\frac{t+1}{t-1}\right) = G_1(t).$$

The equation (8.43) goes over into

$$G_1(-t) = (-1)^q G_1(t),$$

which proves the statement.

Similarly, if we put

$$t = \frac{\tau - \rho}{\tau + \rho^2}, \qquad \tau = -\rho\frac{t\rho + 1}{t-1}, \qquad (\rho = e^{\pi i/3})$$

$$G(\tau) = G\left(-\rho\frac{t\rho+1}{t-1}\right) = G_2(t),$$

we change (8.44) into

$$G_2(-\rho t) = (-\rho)^q G_2(t),$$

which means that in the power series

$$G_2(t) = c_0 + c_1 t + c_2 t^2 + \cdots$$

only such c_n can be different from 0 for which

$$n \equiv q \pmod 3.$$

If therefore $q \not\equiv 0 \pmod 3$, $G(\tau)$ has at $\tau = \rho$ (and because of (8.42) also at $\tau = \rho^2$) a zero of an order congruent to q modulo 3.

If q is divisible by 6, the equations (8.42), (8.43) show that $G(\tau)$ is an absolute modular function. Otherwise we consider the function

$$\varphi_{\beta,\gamma}(\tau) = \sqrt{J(\tau) - 1}^{\beta} \cdot \sqrt[3]{J(\tau)}^{\gamma}, \tag{8.51}$$

$$\begin{aligned} &\beta = 0, 1, && \beta \equiv q \pmod 2, \\ &\gamma = 0, 1, 2, && \gamma \equiv q \pmod 3. \end{aligned} \tag{8.52}$$

Here $J(\tau)$ is the absolute modular invariant, which we normalize as

$$J(\tau) = J\left(\frac{\omega_2}{\omega_1}\right) = \frac{g_2^3}{g_2^3 - 27g_3^2} = \frac{g_2^3(\omega_1, \omega_2)}{\Delta(\omega_1, \omega_2)}, \tag{8.53}$$

where we have

$$\Delta(\omega_1, \omega_2) = g_2^3 - 27g_3^2 = \frac{\pi^{12}}{\omega_1^{12}} \eta(\tau)^{24},$$

$$g_2 = g_2(\omega_1, \omega_2) = 60 \sum{}' \frac{1}{(2m_1\omega_1 + 2m_2\omega_2)^4},$$

$$g_3 = g_3(\omega_1, \omega_2) = 140 \sum{}' \frac{1}{(2m_1\omega_1 + 2m_2\omega_2)^6}.$$

With these expressions we define the roots in (8.51) as

$$\sqrt{J(\tau) - 1} = \frac{3^{\frac{1}{2}}}{\pi^6} \cdot \frac{g_3(1, \tau)}{\eta(\tau)^{12}}, \tag{8.61}$$

$$\sqrt[3]{J(\tau)} = \frac{1}{\pi^4} \cdot \frac{g_2(1, \tau)}{\eta(\tau)^8}. \tag{8.62}$$

For these functions we find easily, by means of (8.21), (8.22),

$$\sqrt{J(\tau + 1) - 1} = -\sqrt{J(\tau) - 1}, \qquad \sqrt{J\left(-\frac{1}{\tau}\right) - 1} = -\sqrt{J(\tau) - 1} \tag{8.63}$$

and

$$\sqrt[3]{J(\tau + 1)} = e^{-2\pi i/3} \sqrt[3]{J(\tau)}, \qquad \sqrt[3]{J\left(-\frac{1}{\tau}\right)} = \sqrt[3]{J(\tau)}, \tag{8.64}$$

so that we have

$$\varphi_{\beta,\gamma}(\tau + 1) = e^{\pi i q/3} \varphi_{\beta,\gamma}(\tau), \tag{8.71}$$

$$\varphi_{\beta,\gamma}\left(-\frac{1}{\tau}\right) = e^{\pi i q} \varphi_{\beta,\gamma}(\tau), \tag{8.72}$$

is always an integer, and we can write

$$G(\tau + 1) = e^{\pi i q/3} G(\tau), \tag{8.42}$$

$$G\left(-\frac{1}{\tau}\right) = e^{\pi i q} G(\tau). \tag{8.43}$$

The function $G(\tau)$ is therefore a modular form of dimension zero, i.e. a "modular function." A consequence of the last two equations is

$$G\left(1 - \frac{1}{\tau}\right) = e^{4\pi i q/3} G(\tau). \tag{8.44}$$

By expanding around the point $\tau = i$ we can see from (8.43) that, for odd q, $G(\tau)$ must have a zero of odd order. For this purpose we put

$$t = \frac{\tau - i}{\tau + i}, \qquad \tau = -i\,\frac{t+1}{t-1},$$

$$G(\tau) = G\left(-i\,\frac{t+1}{t-1}\right) = G_1(t).$$

The equation (8.43) goes over into

$$G_1(-t) = (-1)^q G_1(t),$$

which proves the statement.

Similarly, if we put

$$t = \frac{\tau - \rho}{\tau + \rho^2}, \qquad \tau = -\rho\,\frac{t\rho + 1}{t-1}, \qquad (\rho = e^{\pi i/3})$$

$$G(\tau) = G\left(-\rho\,\frac{t\rho+1}{t-1}\right) = G_2(t),$$

we change (8.44) into

$$G_2(-\rho t) = (-\rho)^q G_2(t),$$

which means that in the power series

$$G_2(t) = c_0 + c_1 t + c_2 t^2 + \cdots$$

only such c_n can be different from 0 for which

$$n \equiv q \pmod 3.$$

If therefore $q \not\equiv 0 \pmod 3$, $G(\tau)$ has at $\tau = \rho$ (and because of (8.42) also at $\tau = \rho^2$) a zero of an order congruent to q modulo 3.

If q is divisible by 6, the equations (8.42), (8.43) show that $G(\tau)$ is an absolute modular function. Otherwise we consider the function

$$\varphi_{\beta,\gamma}(\tau) = \sqrt{J(\tau) - 1}^{\beta} \cdot \sqrt[3]{J(\tau)}^{\gamma}, \tag{8.51}$$

$$\begin{aligned} &\beta = 0, 1, && \beta \equiv q \pmod 2, \\ &\gamma = 0, 1, 2, && \gamma \equiv q \pmod 3. \end{aligned} \tag{8.52}$$

Here $J(\tau)$ is the absolute modular invariant, which we normalize as

$$J(\tau) = J\left(\frac{\omega_2}{\omega_1}\right) = \frac{g_2^3}{g_2^3 - 27g_3^2} = \frac{g_2^3(\omega_1, \omega_2)}{\Delta(\omega_1, \omega_2)}, \tag{8.53}$$

where we have

$$\Delta(\omega_1, \omega_2) = g_2^3 - 27g_3^2 = \frac{\pi^{12}}{\omega_1^{12}} \eta(\tau)^{24},$$

$$g_2 = g_2(\omega_1, \omega_2) = 60 \sum{}' \frac{1}{(2m_1\omega_1 + 2m_2\omega_2)^4},$$

$$g_3 = g_3(\omega_1, \omega_2) = 140 \sum{}' \frac{1}{(2m_1\omega_1 + 2m_2\omega_2)^6}.$$

With these expressions we define the roots in (8.51) as

$$\sqrt{J(\tau) - 1} = \frac{3^{\frac{3}{4}}}{\pi^6} \cdot \frac{g_3(1, \tau)}{\eta(\tau)^{12}}, \tag{8.61}$$

$$\sqrt[3]{J(\tau)} = \frac{1}{\pi^4} \cdot \frac{g_2(1, \tau)}{\eta(\tau)^8}. \tag{8.62}$$

For these functions we find easily, by means of (8.21), (8.22),

$$\sqrt{J(\tau + 1) - 1} = -\sqrt{J(\tau) - 1}, \qquad \sqrt{J\left(-\frac{1}{\tau}\right) - 1} = -\sqrt{J(\tau) - 1} \tag{8.63}$$

and

$$\sqrt[3]{J(\tau + 1)} = e^{-2\pi i/3} \sqrt[3]{J(\tau)}, \qquad \sqrt[3]{J\left(-\frac{1}{\tau}\right)} = \sqrt[3]{J(\tau)}, \tag{8.64}$$

so that we have

$$\varphi_{\beta,\gamma}(\tau + 1) = e^{\pi i q/3} \varphi_{\beta,\gamma}(\tau), \tag{8.71}$$

$$\varphi_{\beta,\gamma}\left(-\frac{1}{\tau}\right) = e^{\pi i q} \varphi_{\beta,\gamma}(\tau), \tag{8.72}$$

since β and γ satisfy the congruences (8.52). The function $\varphi_{\beta,\gamma}(\tau)$ has a zero of order β at $\tau = i$ and a zero of order γ at $\tau = \rho$. The function

$$H(\tau) = \frac{G(\tau)}{\varphi_{\beta,\gamma}(\tau)} \tag{8.81}$$

(which now may include the trivial case $\varphi_{0,0}(\tau) = 1$ for $q \equiv 0 \pmod 6$) is therefore regular in the upper half-plane, and in consequence of (8.42), (8.43), (8.71), (8.72) it fulfills the equations

$$H(\tau + 1) = H(\tau), \qquad H\left(-\frac{1}{\tau}\right) = H(\tau). \tag{8.82}$$

If therefore $H(\tau)$ is not a constant it must be an entire rational function of $J(\tau)$.

Combining this result with the definitions (8.3), (8.81) we find that

$$F(\tau) = \eta(\tau)^{-2r}\varphi_{\beta,\gamma}(\tau)\{B_\kappa J(\tau)^\kappa + \cdots + B_1 J(\tau) + B_0\} \tag{8.91}$$

or, because of (8.51), (8.61), (8.62),

$$F(\tau) = \eta(\tau)^{-2r-12\beta-8\gamma} g_3(1, \tau)^\beta g_2(1, \tau)^\gamma \{C_\kappa J(\tau)^\kappa + \cdots + C_1 J(\tau) + C_0\} \tag{8.92}$$

is the most general function satisfying our postulates (1.22), (1.52), (1.71).

The independent parameters of $F(\tau)$ are r, β, γ, κ, C_0, C_1, $\cdots$ C_κ, which are subject to the restrictions that r is real, β, γ, κ are integers and

$$r > 0; \qquad \beta = 0, 1, \qquad \gamma = 0, 1, 2; \qquad \kappa \geqq 0. \tag{8.93}$$

In order to express the numbers α and μ of (1.71) by these parameters we compute the first term with negative exponent of

$$\eta(\tau)^{-2r}\sqrt{J(\tau) - 1}^{\beta} \cdot \sqrt[3]{J(\tau)}^{\gamma} \cdot J(\tau)^\kappa$$

and compare it with (1.71). Since we have

$$J(\tau) = \frac{e^{-2\pi i\tau}}{3^3 \cdot 2^6}\,(1 + 744e^{2\pi i\tau} + \cdots)$$

and because of (8.11) this comparison yields

$$\frac{r}{12} + \frac{\beta}{2} + \frac{\gamma}{3} + \kappa = \mu - \alpha,$$

$$\mu = \kappa + \frac{r}{12} + \frac{\beta}{2} + \frac{\gamma}{3} + \alpha.$$

This is indeed an integer, since in consequence of (8.52) and (8.41)

$$3\beta + 2\gamma \equiv -6\alpha - r/2 \pmod 6.$$

We see moreover that under the conditions (8.93) always $\mu \geqq 1$, as it has to be. Observing that $0 \leqq \alpha < 1$, we can express α and μ in the following manner

$$\alpha = -\frac{r}{12} - \frac{\beta}{2} - \frac{\gamma}{3} - \left[-\frac{r}{12} - \frac{\beta}{2} - \frac{\gamma}{3}\right], \tag{8.94}$$

$$\mu = \kappa - \left[-\frac{r}{12} - \frac{\beta}{2} - \frac{\gamma}{3}\right], \tag{8.95}$$

where $[x]$ denotes the greatest integer not surpassing x.

The coefficients $a_{-\mu}, \cdots, a_{-1}$ cannot be obtained in this concise manner. They need, in each case, a separate calculation from the first coefficients of $\eta(\tau)^{-1}$ and $J(\tau)$ as the examples below illustrate.

9. The multiplier ϵ

The multiplier ϵ in (1.22) and subsequent equations can, however, be explicitly expressed in terms of the parameters r, β, γ. For this purpose we study the multipliers of

$$\eta(\tau)^{-2r}, \qquad \varphi_{\beta,0}(\tau) = \sqrt{J(\tau) - 1}^{\beta} \qquad \varphi_{0,\gamma}(\tau) = \sqrt[3]{J(\tau)}^{\gamma}$$

separately.

Since r can be any real positive number we have to go back to $\log \eta(\tau)$, using Dedekind's formula. We have[9]

$$\eta\left(\frac{a\tau + b}{c\tau + d}\right)^{-2r} = \epsilon' \cdot (-i(c\tau + d))^{-r} \eta(\tau)^{-2r} \tag{9.11}$$

with

$$\epsilon' = \exp\left\{-2\pi i r\left(\frac{a + d}{12c} - s(a, c)\right)\right\}, \qquad c > 0, \tag{9.12}$$

$$s(a, c) = \sum_{l=1}^{c-1} \frac{l}{c}\left(\frac{al}{c} - \left[\frac{al}{c}\right] - \frac{1}{2}\right).$$

For $\varphi_{1,0}(\tau) = (J(\tau) - 1)^{1/2}$ we make use of (8.63), writing it

$$\varphi_{1,0}(\tau + 1) = -\varphi_{1,0}(\tau), \qquad \varphi_{1,0}(-1/\tau) = -\varphi_{1,0}(\tau). \tag{9.21}$$

It is easily seen that $\varphi_{1,0}(\tau)$ is invariant under the transformations

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \tag{9.22}$$

---

[9] H. Rademacher, *Zur Theorie der Modulfunktionen*, Journal für die reine und angewandte Mathematik, vol. 167 (1931), 312–336, in particular p. 322, formula (3.19).

which are the generators of the principal congruence subgroup mod 2 of index 6 in the complete modular group.[10] A system of representatives of the factor group is given by

$$\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \begin{pmatrix}1&1\\0&1\end{pmatrix},\quad \begin{pmatrix}1&0\\1&1\end{pmatrix},\quad \begin{pmatrix}0&-1\\1&0\end{pmatrix},\quad \begin{pmatrix}0&-1\\1&1\end{pmatrix},\quad \begin{pmatrix}1&-1\\1&0\end{pmatrix}, \tag{9.23}$$

and we find from (9.21)

$$\varphi_{1,0}(\tau)=\varphi_{1,0}(\tau),\quad \varphi_{1,0}(\tau+1)=-\varphi_{1,0}(\tau),\quad \varphi_{1,0}\left(\frac{\tau}{\tau+1}\right)=-\varphi_{1,0}(\tau),$$
$$\varphi_{1,0}\left(-\frac{1}{\tau}\right)=-\varphi_{1,0}(\tau),\quad \varphi_{1,0}\left(\frac{-1}{\tau+1}\right)=\varphi_{1,0}(\tau),\quad \varphi_{1,0}\left(\frac{\tau-1}{\tau}\right)=\varphi_{1,0}(\tau). \tag{9.24}$$

Now all cases (9.24) and hence all modular substitutions satisfy

$$\varphi_{1,0}\left(\frac{a\tau+b}{c\tau+d}\right)=\epsilon''\cdot\varphi_{1,0}(\tau) \tag{9.25}$$

with

$$\epsilon''=\exp\{\pi i(b(a+d)+a(b-c)+bc)\}, \tag{9.26}$$

as seen by a direct verification.

For $\varphi_{0,1}(\tau)=J(\tau)^{1/3}$ we have from (8.64)

$$\varphi_{0,1}(\tau+1)=\xi\cdot\varphi_{0,1}(\tau),\quad \varphi_{0,1}(-1/\tau)=\varphi_{0,1}(\tau),\quad \xi=e^{-2\pi i/3}. \tag{9.31}$$

As a consequence of these equations the function $\varphi_{0,1}(\tau)$ turns out to be invariant with respect to the substitutions

$$\begin{pmatrix}1&3\\0&1\end{pmatrix},\quad \begin{pmatrix}-1&0\\3&-1\end{pmatrix},\quad \begin{pmatrix}2&-3\\3&-4\end{pmatrix},\quad \begin{pmatrix}-4&-3\\3&2\end{pmatrix}, \tag{9.32}$$

which form a system of generators for the principal congruence subgroup modulo 3.[11] As a system of representatives for the factor group the following substitutions can be taken

$$\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \begin{pmatrix}1&1\\0&1\end{pmatrix},\quad \begin{pmatrix}1&-1\\0&1\end{pmatrix},\quad \begin{pmatrix}1&0\\1&1\end{pmatrix},$$
$$\begin{pmatrix}1&1\\1&2\end{pmatrix},\quad \begin{pmatrix}1&-1\\1&0\end{pmatrix},\quad \begin{pmatrix}0&-1\\1&0\end{pmatrix},\quad \begin{pmatrix}0&-1\\1&1\end{pmatrix},$$
$$\begin{pmatrix}0&-1\\1&-1\end{pmatrix},\quad \begin{pmatrix}-1&0\\1&-1\end{pmatrix},\quad \begin{pmatrix}-1&-1\\1&0\end{pmatrix},\quad \begin{pmatrix}-1&1\\1&-2\end{pmatrix},$$

[10] Klein-Fricke, *Vorlesungen über die Theorie der elliptischen Modulfunktionen*, Band 1, p. 276.

[11] Klein-Fricke, l.c., p. 354.

and we find (in the same order), applying (9.31)

$$\varphi_{0,1}(\tau) = \varphi_{0,1}(\tau), \qquad \varphi_{0,1}(\tau+1) = \xi\varphi_{0,1}(\tau), \qquad \varphi_{0,1}(\tau-1) = \xi^{-1}\varphi_{0,1}(\tau),$$

$$\varphi_{0,1}\left(\frac{\tau}{\tau+1}\right) = \xi^{-1}\varphi_{0,1}(\tau), \qquad \varphi_{0,1}\left(\frac{\tau+1}{\tau+2}\right) = \varphi_{0,1}(\tau),$$

$$\varphi_{0,1}\left(\frac{\tau-1}{\tau}\right) = \xi\varphi_{0,1}(\tau), \qquad \varphi_{0,1}\left(-\frac{1}{\tau}\right) = \varphi_{0,1}(\tau), \qquad \varphi_{0,1}\left(\frac{-1}{\tau+1}\right) = \xi\varphi_{0,1}(\tau),$$

$$\varphi_{0,1}\left(\frac{-1}{\tau-1}\right) = \xi^{-1}\varphi_{0,1}(\tau), \qquad \varphi_{0,1}\left(\frac{-\tau}{\tau-1}\right) = \xi\varphi_{0,1}(\tau),$$

$$\varphi_{0,1}\left(\frac{-\tau-1}{\tau}\right) = \xi^{-1}\varphi_{0,1}(\tau), \qquad \varphi_{0,1}\left(\frac{-\tau+1}{\tau-2}\right) = \varphi_{0,1}(\tau).$$

We can verify for these 12 cases directly that

$$\varphi_{0,1}\left(\frac{a\tau+b}{c\tau+d}\right) = \epsilon'''\cdot\varphi_{0,1}(\tau) \tag{9.33}$$

with

$$\epsilon''' = \xi^{-(a+d)(b-c)(ad+bc)} = \exp\{(2\pi i/3)(a+d)(b-c)(ad+bc)\}, \tag{9.34}$$

which, because of the invariance with respect to the transformations (9.32), is then true for all modular transformations. We remember

$$\varphi_{\beta,\gamma}(\tau) = \varphi_{1,0}(\tau)^{\beta}\cdot\varphi_{0,1}(\tau)^{\gamma},$$

and collecting the results from (8.92), (9.11), (9.12), (9.25), (9.26), (9.33), (9.34) we can state the

THEOREM 2: *The modular form of dimension r*

$$F(\tau) = \eta(\tau)^{-2r-12\beta-8\gamma}g_3(1,\tau)^{\beta}g_2(1,\tau)^{\gamma}\{C_\kappa J(\tau)^{\kappa} + \cdots + C_1 J(\tau) + C_0\} \tag{9.41}$$

satisfies the transformation equations

$$F\left(\frac{a\tau+b}{c\tau+d}\right) = \epsilon(a,b,c,d)\cdot(-i(c\tau+d))^{-r}F(\tau), \qquad c>0, \tag{9.42}$$

and

$$F(\tau+1) = e^{2\pi i a}F(\tau), \tag{9.43}$$

with the multiplier

$$\epsilon(a,b,c,d) = \exp\left\{2\pi i\left(rs(a,c) - r\frac{a+d}{12c} + \frac{\beta}{2}(b(a+d)+a(b-c)+bc) + \frac{\gamma}{3}(a+d)(b-c)(ad+bc)\right)\right\}, \tag{9.44}$$

[12]

[12] It can easily be proved that the multiplier ϵ defined by (9.44) fulfills, because of (9.45), the condition mentioned in footnote 4.

and with

$$\alpha = -\frac{r}{12} - \frac{\beta}{2} - \frac{\gamma}{3} - \left[-\frac{r}{12} - \frac{\beta}{2} - \frac{\gamma}{3}\right]. \tag{9.45}$$

Conversely, all modular forms regular in the upper half-plane, with at most a polar singularity at $i\infty$, and satisfying (9.42), (9.43), *are contained in* (9.41).

We finally apply (9.44) to obtain an expression for $A_{k,\nu}(m)$ of (5.6). By a straightforward calculation making use of

$$s(h', k) = -s(h, k),^{13}$$

we obtain for $A_{k,\nu}(m)$ the formulae (9.53), (9.54) of the following

THEOREM 3: *The modular form* (9.41) *of positive dimension r admits for $\Im(\tau) > 0$ the Fourier expansion*

$$F(\tau) = e^{2\pi i a\tau}\left\{a_{-\mu}e^{-2\pi i\mu\tau} + \cdots + a_{-1}e^{-2\pi i\tau} + \sum_{m=0}^{\infty} a_m e^{2\pi i m\tau}\right\} \tag{9.51}$$

with the coefficients

$$a_m = 2\pi \sum_{\nu=1}^{\mu} a_{-\nu} \sum_{k=1}^{\infty} \frac{1}{k} A_{k,\nu}(m)\left(\frac{\nu - \alpha}{m + \alpha}\right)^{(r+1)/2} I_{r+1}((4\pi/k)(\nu - \alpha)^{\frac{1}{2}}(m + \alpha)^{\frac{1}{2}}), \tag{9.52}$$

for $m \geqq 0$, where the number μ is determined by (8.95) *and where*

$$A_{k,\nu}(m) = \sum_{\substack{0 \leqq h < k \\ (h,k)=1}} \omega_r(h, k)\cdot\xi_1(h, k)^{\beta}\cdot\xi_2(h, k)^{\gamma} \times \exp\left[-(2\pi i/k)((\nu - \mu + \kappa)h' + (m + \mu - \kappa)h)\right] \tag{9.53}$$

with

$$\omega_r(h, k) = \exp\{2\pi i r s(h, k)\},$$

$$\xi_1(h, k) = \exp\left\{\pi i\left(-\frac{h'(h^2 + 1)}{k} + h'k + hh' + 1\right)\right\}, \tag{9.54}$$

$$\xi_2(h, k) = \exp\left\{-\frac{2\pi i}{3}(h - h')\left(\left(\frac{hh' + 1}{k} + k\right)(2hh' + 1) + \frac{1}{k}\right)\right\}.$$

It is of interest to notice that the real number r enters only into $\omega_r(h, k)$, whereas the three other factors of the summands in (9.53) are always roots of unity.

10. Examples

The following examples may show how in each case the coefficients $a_{-\mu}, \cdots, a_{-1}$ of the principal part at $\tau = i\infty$ can be directly computed. As far as the selection of the examples is concerned, we have restricted ourselves

---

[13] Rademacher, l.c. footnote 9, formula (2.13).

to functions which lead to not too complicated coefficients $A_{k,\nu}(m)$. In some examples we get the same coefficients which appeared first in the Hardy-Ramanujan formula. We shall give the formulae without further explanations, since they can be understood as simple applications of Theorems 2 and 3.

I. $r = \frac{1}{2}, \quad \beta = \gamma = \kappa = 0$:

$$F(\tau) = \eta(\tau)^{-1},$$

$$\alpha = -1/24 - [-1/24] = 23/24,$$

$$\mu = 0 - [-1/24] = 1.$$

$$\eta(\tau)^{-1} = e^{23\pi i\tau/12} e^{-2\pi i\tau} \prod_{l=1}^{\infty} (1 - e^{2\pi i l\tau})^{-1} = e^{23\pi i\tau/12} \cdot \left\{ e^{-2\pi i\tau} + \sum_{m=0}^{\infty} a_m e^{2\pi i m\tau} \right\}. \tag{10.11}$$

Hence:

$$a_{-1} = 1,$$

and

$$a_m = 2\pi \sum_{k=1}^{\infty} \frac{1}{k} A_{k,1}(m) \left(\frac{\frac{1}{24}}{m + \frac{23}{24}} \right)^{\frac{3}{4}} I_{\frac{1}{2}} \left(\frac{4\pi}{k} \left[\frac{1}{24} \left(m + \frac{23}{24} \right) \right]^{\frac{1}{2}} \right) \tag{10.12}$$

with

$$A_{k,1}(m) = \sum_{\substack{0 \le h < k \\ (h,k)=1}} \omega_{\frac{1}{2}}(h, k)\, e^{-2\pi i(m+1)h/k},$$

$$\omega_{\frac{1}{2}}(h, k) = \exp\{\pi i\, s(h, k)\} = \omega_{h,k}, \tag{10.13}$$

$$A_{k,1}(m) = A_k(m + 1)$$

where $\omega_{h,k}$ and $A_k(r)$ are the notations used by Hardy and Ramanujan.[14] Now we have

$$\frac{d}{dm} \left(\frac{I_\rho(A(m+\alpha)^{\frac{1}{2}})}{(m+\alpha)^{\rho/2}} \right) = \frac{A}{2} \frac{I_{\rho+1}(A(m+\alpha)^{\frac{1}{2}})}{(m+\alpha)^{(\rho+1)/2}}.^{15} \tag{10.14}$$

Since moreover

$$I_{\frac{1}{2}}(z) = (2/\pi z)^{\frac{1}{2}} \sinh z,^{16} \tag{10.15}$$

---

[14] Proc. London Math. Soc. (2) vol. 17 (1918), 75–115, in particular formulae (1.721), (1.722), (1.73); cf. Rademacher, Journal London Math. Soc. 7 (1932), 14–19, formula (5).

[15] An immediate consequence of formula (6), in Watson, l.c., p. 79.

[16] Watson, l.c., p. 80, (10).

the equation (10.12) yields, after some simplifications,

$$a_m = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(m+1)k^{\frac{1}{2}} \frac{d}{dm}\left(\frac{\sinh\left(\frac{4\pi}{k}\left[\frac{1}{24}\left(m+\frac{23}{24}\right)\right]^{\frac{1}{2}}\right)}{\left[m+\frac{23}{24}\right]^{\frac{1}{2}}}\right).$$

Taking note of the remark in footnote[3] and putting $m+1=n$, we have again

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n)k^{\frac{1}{2}} \frac{d}{dn}\left(\frac{\sinh\left(\frac{\pi}{k}\left[\frac{2}{3}\left(n-\frac{1}{24}\right)\right]^{\frac{1}{2}}\right)}{\left[n-\frac{1}{24}\right]^{\frac{1}{2}}}\right), \tag{10.16}$$

as proved in a previous paper.

II. $r = 12, \quad \beta = \gamma = \kappa = 0$:

$$F(\tau) = \eta(\tau)^{-24} = \pi^{12}\Delta(1,\tau)^{-1},$$

$$\alpha = -1 - [-1] = 0, \tag{10.21}$$

$$\mu = 0 - [-1] = 1.$$

Now

$$\begin{aligned}\eta(\tau)^{-24} &= e^{-2\pi i\tau}\prod_{l=1}^{\infty}(1-e^{2\pi il\tau})^{-24}\\ &= e^{-2\pi i\tau}\{1+e^{2\pi i\tau}+\cdots\}^{24} \qquad (10.22)\\ &= e^{-2\pi i\tau}+\sum_{m=0}^{\infty}a_m e^{2\pi im\tau},\end{aligned}$$

from which we infer

$$a_{-1} = 1.$$

Therefore

$$a_m = 2\pi\sum_{k=1}^{\infty}\frac{1}{k}A_{k,1}(m)\,m^{-13/2}I_{13}(4\pi m^{\frac{1}{2}}k^{-1}). \tag{10.23}$$

Here we have

$$A_{k,1}(m) = \sum_{\substack{h \bmod k\\(h,k)=1}} \omega_{12}(h,k)\,e^{-2\pi i(m+1)h/k},$$

$$\omega_{12}(h,k) = \exp\{2\pi i\cdot 12s(h,k)\}.$$

Since

$$12s(h,k) \equiv \frac{h-h'}{k} \pmod 1, \tag{10.24}$$

[17] Rademacher, l.c.[9], p. 321, formula (2.51).

we have

$$\omega_{12}(h, k) = e^{2\pi i(h-h')/k}$$

and therefore

$$A_{k,1}(m) = \sum_{\substack{h \bmod k \\ (h,k)=1}} e^{-2\pi i(h'+mh)/k}, \tag{10.25}$$

Thus

$$a_m = 2\pi \sum_{k=1}^{\infty} \frac{1}{k} \sum_{\substack{h \bmod k \\ (h,k)=1}} e^{-2\pi i(h'+mh)/k} m^{-13/2} I_{13}(4\pi m^{\frac{1}{2}} k^{-1}). \tag{10.26}$$

This formula permits of an explicit determination of a_0 and so leads to a certain verification, since (10.22) furnishes directly $a_0 = 24$. On the other hand we have

$$A_{k,1}(0) = \sum_{\substack{h \bmod k \\ (h,k)=1}} e^{-2\pi i h'/k} = \mu(k),$$

$\mu(k)$ designating the Möbius function, and according to (6.2) we are to take the interpretation

$$m^{-13/2} I_{13}(4\pi m^{\frac{1}{2}} k^{-1})_{m=0} = \frac{(2\pi)^{13}}{k^{13} \cdot 13!}.$$

Therefore (10.23) yields

$$a_0 = \frac{(2\pi)^{14}}{13!} \sum_{k=1}^{\infty} \frac{\mu(k)}{k^{14}} = \frac{(2\pi)^{14}}{13!} \frac{1}{\zeta(14)} = 28/B_7 = 24,$$

in accordance with the previously obtained value.

III. $\qquad 12 < r \leqq 24, \qquad \beta = \gamma = \kappa = 0.$

$$\alpha = -r/12 - [-r/12] = 2 - r/12, \qquad \mu = 0 - [-r/12] = 2,$$

$$F(\tau) = \eta(\tau)^{-2r} = e^{2\pi i r(2-r/12)} \left\{ e^{-4\pi i \tau} + 2re^{-2\pi i \tau} + \sum_{m=0}^{\infty} a_m e^{2\pi i m \tau} \right\}.$$

Therefore

$$a_{-2} = 1, \qquad a_{-1} = 2r.$$

Let us consider e.g. the special case $r = 25/2$:

$$F(\tau) = \eta(\tau)^{-25} = e^{23\pi i \tau/12} \left\{ e^{-4\pi i \tau} + 25e^{-2\pi i \tau} + \sum_{m=0}^{\infty} a_m e^{2\pi i m \tau} \right\}, \tag{10.31}$$

$$a_{-2} = 1, \qquad a_{-1} = 25. \tag{10.32}$$

We find by means of (10.24) and (10.13)

$$A_{k,1}(m) = \sum_{\substack{h \bmod k \\ (h,k)=1}} \omega_{h,k} e^{-2\pi i(m+1)h/k}, \tag{10.33}$$

$$A_{k,2}(m) = \sum_{\substack{h \bmod k \\ (h,k)=1}} \omega_{h,k} e^{-2\pi i(h'+(m+1)h)/k}, \tag{10.34}$$

$$a_m = 2\pi \left\{ 25 \sum_{k=1}^{\infty} \frac{1}{k} A_{k,1}(m) \left(\frac{\frac{1}{24}}{m+\frac{23}{24}} \right)^{27/4} I_{27/2}\left(\frac{4\pi}{k}\left[\frac{1}{24}\left(m+\frac{23}{24} \right) \right]^{\frac{1}{2}} \right) \right.$$

$$\left. + \sum_{k=1}^{\infty} \frac{1}{k} A_{k,2}(m) \left(\frac{\frac{25}{24}}{m+\frac{23}{24}} \right)^{27/4} I_{27/2}\left(\frac{4\pi}{k}\left[\frac{25}{24}\left(m+\frac{23}{24} \right) \right]^{\frac{1}{2}} \right) \right\}. \tag{10.35}$$

If we write, in order to exhibit the linearity in the coefficients a_{-2} and a_{-1}

$$F(\tau) = a_{-2} e^{23\pi i\tau/12} \left\{ e^{-4\pi i\tau} + 2\pi \sum_{m=0}^{\infty} e^{2\pi i m\tau} \sum_{k=1}^{\infty} \frac{1}{k} A_{k,2}(m)\, 5^{27/2} \frac{I_{27/2}\left(\frac{5\pi}{k}\left[\frac{2}{3}\left(m+\frac{23}{24} \right) \right]^{\frac{1}{2}} \right)}{(24m+23)^{27/4}} \right\}$$

$$+ a_{-1} e^{23\pi i\tau/12} \left\{ e^{-2\pi i\tau} + 2\pi \sum_{m=0}^{\infty} e^{2\pi i m\tau} \sum_{k=1}^{\infty} \frac{1}{k} A_{k,1}(m) \frac{I_{27/2}\left(\frac{\pi}{k}\left[\frac{2}{3}\left(m+\frac{23}{24} \right) \right]^{\frac{1}{2}} \right)}{(24m+23)^{27/4}} \right\} \tag{10.36}$$

$$= a_{-2} F_2(\tau) + a_{-1} F_1(\tau),$$

we see that a modular form arises only for the ratio

$$a_{-2} : a_{-1} = 1 : 25. \tag{10.37}$$

Indeed, no other linear combination of $F_2(\tau)$ and $F_1(\tau)$, nor $F_2(\tau)$ or $F_1(\tau)$ separately can be a modular form (at least not for the full modular group). A comparison with (9.52) shows that any such combination would belong to

$$\alpha = \frac{23}{24}, \qquad r = \frac{25}{2}.$$ [18]

---

[18] This argument might need some further explanation. The question is, whether, given a function of type (9.51), the values α, r and μ in (9.52) are (if at all) *uniquely* determined. This is, indeed, the case. The value of α with $0 \leqq \alpha < 1$ can be derived from the fact that $F(\tau) \exp(-2\pi i\alpha\tau)$ is periodic in τ mod 1. The value for r appears in the form $(r+1)/2$ as exponent in (9.52), and since

$$I_{r+1}(4\pi(\mu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}) \sim \frac{\exp\left[4\pi(\mu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}\right]}{2\pi 2^{\frac{1}{2}}(\mu-\alpha)^{\frac{1}{4}}(m+\alpha)^{\frac{1}{4}}}$$

we have

$$a_m \sim C \frac{\exp\left[4\pi(\mu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}}\right]}{(m+\alpha)^{r/2+3/4}}.$$

Hence we obtain

The diophantine equation (9.45) for β and γ would become

$$\frac{23}{24} \equiv -\frac{25}{24} - \frac{\beta}{2} - \frac{\gamma}{3} \pmod 1,$$

which yields, among the admissible values of β and γ, only

$$\beta = 0, \qquad \gamma = 0.$$

Moreover we should have in general $\mu = 2$, and $\mu = 1$ only for the case $a_{-2} = 0$. We obtain from (8.95)

$$2 \geqq \kappa - [-25/24] = \kappa + 2$$

which together with $\kappa \geqq 0$ leads to the determination

$$\kappa = 0.$$

Our reasoning shows that any linear combination

$$a_{-2}F_2(\tau) + a_{-1}F_1(\tau) \tag{10.38}$$

which is a modular form has the uniquely determined values for r, β, γ, r which we just have computed. But these numbers in turn determine, by (9.41) of Theorem 2, uniquely (up to a constant factor) the modular form to which they belong. Thus (10.31) is the only modular form contained in the linear set (10.38), and therefore the condition (10.37) is necessary.

This discussion shows that in general the coefficients of the principal part (1.8) are not independent of each other. This is not surprising, since it follows from (8.95) that

$$\mu = \kappa + 1$$

occurs only for

$$\frac{r}{12} + \frac{\beta}{2} + \frac{\gamma}{3} \leqq 1.$$

Otherwise we have

$$\mu > \kappa + 1,$$

which means that there are less independent parameters $C_0, C_1, \cdots C_\kappa$ than coefficients $a_{-1}, \cdots, a_{-\mu}$.

$$\mu - \alpha = \lim_{m\to\infty} \frac{(\log a_m)^2}{16\pi^2 m}$$

and then

$$\frac{r}{2} + \frac{3}{4} = \lim_{m\to\infty} \frac{4\pi(\mu-\alpha)^{\frac{1}{2}}(m+\alpha)^{\frac{1}{2}} - \log a_m}{\log m}$$

as *necessary* conditions for μ and r.

IV. All our preceding examples assumed $\kappa = 0$. We give a last example with $\kappa > 0$:

$$r = \tfrac{1}{2}, \qquad \beta = \gamma = 0, \qquad \kappa = 1:$$

$$F(\tau) = \eta(\tau)^{-1}\{3^3 \cdot 2^6 J(\tau) + C_0\} \tag{10.41}$$

with arbitrary C_0. We find

$$\alpha = 23/24, \qquad \mu = 2$$

and have

$$F(\tau) = e^{23\pi i\tau/12}\{e^{-2\pi i\tau} + 1 + \cdots\}\{e^{-2\pi i\tau} + (744 + C_0) + \cdots\}$$

$$= e^{23\pi i\tau/12}\left\{e^{-4\pi i\tau} + (745 + C_0)e^{-2\pi i\tau} + \sum_{m=0}^{\infty} a_m e^{2\pi i m\tau}\right\},$$

and hence

$$a_{-2} = 1, \qquad a_{-1} = 745 + C_0.$$

Therefore

$$a_m = 2\pi(745 + C_0)\sum_{k=1}^{\infty}\frac{1}{k}A_{k,1}(m)\left(\frac{\frac{1}{24}}{m + \frac{23}{24}}\right)^{3/4} I_{3/2}\left(\frac{4\pi}{k}\left[\frac{1}{24}\left(m + \frac{23}{24}\right)\right]^{\frac{1}{2}}\right)$$

$$+ 2\pi\sum_{k=1}^{\infty}\frac{1}{k}A_{k,2}(m)\left(\frac{\frac{25}{24}}{m + \frac{23}{24}}\right)^{3/4} I_{3/2}\left(\frac{4\pi}{k}\left[\frac{25}{24}\left(m + \frac{23}{24}\right)\right]^{\frac{1}{2}}\right), \tag{10.42}$$

where $A_{k,1}(m)$ and $A_{k,2}(m)$ turn out to be the numbers (10.33), (10.34) of the preceding example ($A_{k,\nu}(m)$ is, as function of r, periodic modulo 12). If we make use of example I, we obtain

$$a_m = (745 + C_0)p(m + 1) + a_m^* \tag{10.43}$$

with

$$a_m^* = \frac{1}{\pi\sqrt{2}}\sum_{k=1}^{\infty} A_{k,2}(m)\, k^{\frac{1}{2}}\frac{d}{dm}\left(\frac{\sinh\left(\frac{5\pi}{k}\left[\frac{2}{3}\left(m + \frac{23}{24}\right)\right]^{\frac{1}{2}}\right)}{\left[m + \frac{23}{24}\right]^{\frac{1}{2}}}\right). \tag{10.44}$$

The function $J(\tau)$, being of dimension zero, does not fall under the scope of our theory. Nevertheless the numbers a_m^* can be used to express rather simply the coefficients of $J(\tau)$. Indeed, we have, with $C_0 = 0$:

$$3^3 \cdot 2^6 \cdot J(\tau) = F(\tau)\eta(\tau)$$
$$= e^{23\pi i\tau/12}\left\{e^{-4\pi i\tau} + \sum_{m=0}^{\infty} a_m^* e^{2\pi i m\tau}\right\} \cdot e^{\pi i\tau/12} \prod_{l=1}^{\infty} (1 - e^{2\pi i l\tau})$$
$$+ 745\, e^{23\pi i\tau/12}\left\{e^{-2\pi i\tau} + \sum_{m=0}^{\infty} p(m+1)\, e^{2\pi i m\tau}\right\} \cdot \eta(\tau)$$
$$= \left\{e^{-2\pi i\tau} + \sum_{m=0}^{\infty} a_m^* e^{2\pi i(m+1)\tau}\right\} \cdot \sum_{\lambda=-\infty}^{+\infty} (-1)^\lambda e^{2\pi i\tau[\lambda(3\lambda-1)/2]} + 745.$$

If we put

$$3^3 \cdot 2^6 \cdot J(\tau) = e^{-2\pi i\tau} + 744 + \sum_{n=1}^{\infty} c_n e^{2\pi i n\tau} \tag{10.45}$$

we obtain for the desired coefficients c_n the finite sum

$$c_n = \sum_{[\lambda(3\lambda-1)/2] \leqq n-1} (-1)^\lambda a^*_{n-1-[\lambda(3\lambda-1)/2]} + \epsilon_n, \tag{10.46}$$

where $\epsilon_n = 0$, if $n+1$ is not a pentagonal number, and $\epsilon_n = (-1)^\rho$ if $n+1 = \rho(3\rho-1)/2$.

11. Concluding remarks

We have discussed modular forms of positive real dimension r, generalizing in this respect the customary notion of modular forms, which admits only integers or half integers as dimensions.[19] But we can even go one step further. Nothing prevents us from introducing complex values for r in the developments of nos. 8 and 9 up to Theorem 2, inclusive. We should in (8.41) again obtain q as an integer, and α would now become a complex number.

This, however, would have the consequence that the multiplier ϵ, which appears in its simplest form in

$$F(\tau+1) = e^{2\pi i\alpha} F(\tau)$$

loses the essential property $|\epsilon| = 1$. This is not an objection against the generalization of the dimension to complex values, but it would completely upset the estimations in nos. 3 and 4, in which $|\Omega_{h,k}| = 1$, or at least $\Omega_{h,k} = 0(1)$ is indispensable. Thus our theory is intrinsically restricted to *real* and positive dimensions of the modular forms.

It will be noticed that among our examples of no. 10 we have not mentioned certain functions which Hardy and Ramanujan discuss incompletely at the end of their paper. But those functions are not accessible by our present method since they either do not belong to the full modular group, but only to a certain congruence subgroup, or they contain functions of dimension zero.

---

[19] See, however, papers of H. Petersson, Mathem. Annalen *103* (1930), Acta Mathem. *58* (1932), Jahresber. Deutsche Math. Ver. *47* (1937).

The first obstacle can be overcome by an appropriate generalization of our method.[20] The second, however, leads to a problem interesting in itself. The series (7.2), in fact, would remain convergent also for $r = 0$, but a direct limit process $r \to 0$ is not obvious. However, the estimates of nos. 3 and 4 can be refined in such a way as to lead to the desired result.[21]

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA.

---

[20] Cf. H. S. Zuckerman, *On the coefficients of certain modular forms belonging to subgroups of the modular group*, to be published in the Transactions of the Amer. Math. Soc.

[21] Cf. H. Rademacher, *The Fourier coefficients of the modular invariant* $J(\tau)$, Amer. Journal of Math., vol. 60, pp. 501–512 (1938).

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

ARC LENGTH IN METRIC AND FINSLER MANIFOLDS

BY SUMNER BYRON MYERS

(Received November 17, 1937)

1. Introduction

We are concerned here with n-dimensional differentiable metric manifolds (locally euclidean metric spaces). An important class of such manifolds consists of Finsler manifolds, in which the metric arises from arc length defined by a positive, positive-regular calculus of variations integral. These include Riemannian manifolds. A well-known property of euclidean space concerning the equivalence of the notion of rectifiable arc and the notion of an arc defined by functions of bounded variation is generalized to a class of metric manifolds including Finsler manifolds. The equivalence of arc length as defined by an integral and arc length as defined by a limit of a sum of lengths of inscribed geodesic arcs holds on Finsler manifolds, and it is further proved that for this purpose geodesic arcs can be replaced by any set of arcs having certain path-properties in common with geodesics. It is shown that if the geodesics of a metric manifold possess these path-properties, the manifold is a Finsler manifold.

2. Let M be a metric space, with a distance function $\rho(P, Q)$ satisfying the usual axioms including the triangle inequality, but not necessarily symmetric. Let c be a directed arc on M, joining a point A to a point B. Consider a set of points $AP_1P_2 \cdots P_mB$ on c, ordered in the direction of a parameter increasing from A to B. Then the least upper bound l of the sum $\sum_{k=0}^{m} (P_k, P_{k+1})$, $(P_0 = A, P_{m+1} = B)$, for all such ordered sets of points and for all m is called the *length* of the directed arc c. If l is finite, c is said to be *rectifiable*. It can be shown that if c is the sum of two arcs, a and b, then the length of c is the sum of the lengths of a and b; also, the length of a subarc a of c varies continuously with the endpoints of a on c. A *geodesic arc* of M is defined to be an arc whose length equals the distance between its endpoints. Note that an arc may be a geodesic if taken in one direction without being a geodesic when taken in the opposite direction. In the future an *arc* will always mean an oriented arc. The length of a geodesic arc is at least as small as the length of any other arc joining its initial point to its final point. Every subarc of a geodesic arc is a geodesic arc. Any triple of points A, B, C on a geodesic arc c ordered in the direction from the initial to the final point of c has the property that B is *between* A and C; that is, $\rho(A, B) + \rho(B, C) = \rho(A, C)$. Conversely, if an arc has this property for all such ordered triples of points it is a geodesic arc.

An important subclass of the metric spaces is formed by the *n-dimensional*

metric manifolds. These are ordinary n-dimensional manifolds (compact or not) metrized so that the topology provided by the metric is equivalent to the original manifold-topology. In these spaces local coördinate systems can be used. If the manifold can be covered by a set of coördinate systems such that in any region of overlapping of two coördinate systems the transformation from one coördinate system to the other is non-singular and of class $C^{(n)}$, we say that the manifold is a *metric manifold of class* $C^{(n)}$. *Admissible* coördinate systems are those obtainable from the given set of coördinate systems by non-singular transformations of class $C^{(n)}$.

A *Finsler manifold* is an n-dimensional manifold of class $C^{(4)}$ in which to each admissible coördinate system (x) is attached a function

$$F(x^1, x^2, \cdots, x^n, r^1, r^2, \cdots, r^n)$$

which is to be invariant under admissible coördinate transformations if (r) is transformed as a contravariant vector. $F(x, r)$ is to be of class $C^{(3)}$ for all (x) and all $(r) \neq (0)$, positive homogeneous of first degree in (r), and is to satisfy the conditions

$$F(x, r) > 0$$

$$| F_{r^i r^j} | \text{ of rank } n - 1$$

for all (x), and all $(r) \neq (0)$. Together these conditions imply the condition of positive regularity,

$$F_{r^i r^j} \lambda_i \lambda_j > 0$$

for all (r) not zero or proportional to (λ). From the homogeneity of F follow the useful relations

$$\begin{aligned} F(x, r) &= F_{r^i} r^i \\ F_{r^i}(x, r) &= F_{r^i}(x, kr) \end{aligned} \qquad (\text{for } k > 0).$$

All these conditions and relations are invariant.

Arc length in a Finsler manifold is given by $\int F(x, x')\, dt$, the integral being the Lebesgue integral along an absolutely continuous representation of a curve. This arc length will be called *integral* arc length, in order to distinguish it from arc length as previously defined for metric spaces, which we shall call *metric* arc length. Short segments of extremals of the integral furnish a proper absolute minimum to the integral arc length of arcs joining their end points. A Finsler manifold is known to become a metric manifold if the distance $\rho(P, Q)$ is defined as the greatest lower bound of the integral arc length of arcs joining P to Q. On the basis of this definition of distance, metric arc length and geodesic arcs may be defined. Along extremals, the two kinds of arc length are equal, so that short extremal segments are geodesic arcs. Conversely, it can be shown that geodesic arcs are extremal segments.

The extremals of a Finsler manifold form a $(2n - 2)$-parameter family, one

through each point in each direction. In a local coördinate system (x), the family can be expressed by functions of class $C^{(2)}$:

$$x^i = x^i(s, \bar{x}^1, \cdots, \bar{x}^n, p^1, \cdots, p^n) \qquad (i = 1, \cdots, n) \tag{1}$$

where

$$\bar{x}^i \equiv x^i(0, \bar{x}, p)$$

$$p^i \equiv x^i_s(0, \bar{x}, p)$$

and s is arc length, so that

$$F(\bar{x}, p) = 1.$$

These functions (1) can be expressed in the form

$$x^i = X^i(\bar{x}^1, \cdots, \bar{x}^n, y^1, \cdots, y^n) \qquad (i = 1, \cdots, n) \tag{2}$$

where

$$y^i = p^i s \tag{3}$$

and

$$\bar{x}^i = X^i(\bar{x}, 0)$$

$$| X^i_{y^j}(x, 0) | \equiv \delta^i_j .$$

The functions X^i are of class C^1, and of class $C^{(2)}$ except at $(y) = (0)$. For a fixed $(\bar{x})$ there is a neighborhood of $(\bar{x})$ in which the x's are in $1 - 1$ correspondence with the y's in a neighborhood $| y^i | < \delta$, with $| X^i_{y^j}(\bar{x}, y) | \neq 0$. The y's are called *normal coördinates with* $(\bar{x})$ *as origin*. There is a neighborhood of every point P which is simply covered by the extremals issuing from P, in which these extremals are geodesic arcs, and in which normal coördinates with P as origin can be used. Such a neighborhood we shall call an *N-neighborhood* of P. Furthermore, every point has a neighborhood in which any two points can be joined by a unique extremal[1] and in which there is no pair of conjugate points, so that this neighborhood is an N-neighborhood of each of its points. Hence for a compact set S there exists an $\epsilon > 0$ independent of P, Q on S such that if $\rho(P, Q) < \epsilon$, P can be joined to Q by a unique geodesic arc and Q lies in an N-neighborhood of P.

3. As is well-known, in euclidean space a necessary and sufficient condition that an arc $x^i = x^i(t)$ $a \leqq t \leqq b$ be rectifiable is that the functions $x^i(t)$ be of bounded variation in the interval $a \leqq t \leqq b$. We shall see that a generalization of this to metric manifolds of class C' can be obtained provided that the manifold M has the following property, which we shall call property I.

Corresponding to an arbitrary compact subset S of M and an arbitrary admissible

[1] J. H. C. Whitehead, Quarterly Journal of Math., vol. 4 (1933), pp. 291–296.

coördinate system (x) *covering it, there exist positive constants* A *and* B *such that for any two points* $P:(x)$ *and* $Q:(\bar{x})$ *of* S

$$(I) \qquad A\,|\,\bar{x}^i - x^i\,| \leqq \rho(P, Q) \leqq B \sum_{j=1}^{n} |\,\bar{x}^j - x^j\,| \qquad (i = 1, \cdots, n).$$

This is really a condition only on points close together, the constants automatically existing for all pairs of points of S whose distance apart is greater than a fixed number d. For corresponding to any constant $d > 0$, there exists a $D > 0$ such that for any two points $P:(x)$ and $Q:(\bar{x})$ on the compact set S for which $\rho(P, Q) > d$, we have $|\,\bar{x}^i - x^i\,| > D$. Let e be the maximum of $\rho(P, Q)$ for all P, Q on S, and let E be the maximum of $|\,\bar{x}^i - x^i\,|$ for all i and all $(\bar{x})$, (x) on S. Then the inequalities (I) are satisfied for all P, Q for which $\rho(P, Q) > d$ if we take

$$A = d/E, \qquad B = e/nD.$$

Property I holds if there is even *one* coördinate system (x) covering S for which constants A and B exist satisfying the inequalities I. For let (y) be any admissible coördinate system different from (x) covering S, and let P and Q be two neighboring points of S, whose coördinates are, in one coördinate system, (x) and $(\bar{x})$, and in the other system (y) and $(\bar{y})$. Since the transformation $y^i = y^i(x)$ and its inverse are of class C', it can be shown that there exists a $d > 0$, an $M > 0$, and an $N > 0$, such that if $\rho(P, Q) < d$,

$$(4) \qquad |\,\bar{y}^i - y^i\,| \leqq M \sum_{j=1}^{n} |\,\bar{x}^j - x^j\,| \qquad (i = 1, \cdots, n)$$

and

$$(5) \qquad |\,\bar{x}^j - x^j\,| \leqq N \sum_{k=1}^{n} |\,\bar{y}^k - y^k\,| \qquad (j = 1, \cdots, n).$$

Now it is given that

$$A\,|\,\bar{x}^i - x^i\,| \leqq \rho(P, Q) \qquad (i = 1, \cdots, n).$$

Hence

$$M \sum_{j=1}^{n} |\,\bar{x}^j - x^j\,| \leqq \frac{Mn}{A}\,\rho(P, Q).$$

Combining this with (4), we obtain

$$|\,\bar{y}^i - y^i\,| \leqq \frac{Mn}{A}\,\rho(P, Q)$$

or

$$\frac{A}{Mn}\,|\,\bar{y}^i - y^i\,| \leqq \rho(P, Q).$$

Furthermore it is given that

$$\rho(P, Q) \leqq B \sum_{j=1}^{n} |\bar{x}^j - x^j|.$$

Combining this with (5), we obtain

$$\rho(P, Q) \leqq BNn \sum_{k=1}^{n} |\bar{y}^k - y^k|.$$

Thus, for $\rho(P, Q) < d$,

$$\frac{A}{Mn} |\bar{y}^i - y^i| \leqq \rho(P, Q) \leqq BNn \sum_{k=1}^{n} |\bar{y}^k - y^k|,$$

which completes the proof.

We shall now prove that *every Finsler manifold has property I.*

Let S be any compact set on the Finsler manifold, (x) any admissible coördinate system covering S, $P:(\bar{x})$ an arbitrary point of S, and (y) normal coördinates with P as origin. Then the transformations

$$x^i = X^i(\bar{x}, y) \qquad y^i = Y^i(\bar{x}, x) \tag{6}$$

are of class C' (even at $(y) = 0$, $(x) = (\bar{x})$). Let Q be any point such that $\rho(P, Q) < \epsilon$; here ϵ is the ϵ of the last statement of §2 taken for our compact set S. The coördinates of Q in the system (y) are given by

$$y^i = p^i s \tag{7}$$

where p^i are the components in the coördinate system (x) of the unit vector at $P:(\bar{x})$ tangent to the geodesic arc joining P to Q and s is the distance from P to Q. Furthermore,

$$1 = F(\bar{x}, p)$$

so that

$$s = F(\bar{x}, ps) = F(\bar{x}, y) = y^i F_{r^i}(\bar{x}, y) = y^i F_{r^i}(\bar{x}, p). \tag{8}$$

For a fixed P, let A be the minimum of $|1/p^i|$ and let B be the maximum of $F_{r^i}(\bar{x}, p)$, for all i and all unit vectors (p) at P. Then it follows from (7) and (8) that

$$Ay^i \leqq s \leqq B \sum_{j=1}^{n} y^j \qquad (i = 1, \cdots, n). \tag{9}$$

If we now allow $P:(\bar{x})$ to vary over the compact set S, and let A be the minimum of $|1/p^i|$ and B the maximum of $F_{r^i}(\bar{x}, p)$ for all i and all unit vectors (p) at all points $(\bar{x})$ of S, (9) will still hold.

But since (6) are of class C', there exists an $M > 0$ such that

$$|x^i - \bar{x}^i| \leqq M \sum_{j=1}^{n} |y^j|$$

and an $N > 0$ such that

$$|y^j| \leqq N \sum_{k=1}^{n} |x^k - \bar{x}^k|$$

where here $\rho(P, Q)$ may have to be restricted to be less than a constant $\eta > 0$ perhaps smaller than ϵ. Combined with (9), these two inequalities yield

$$\frac{A}{Mn} |x^i - \bar{x}^i| \leqq s \leqq BNn \sum_{k=1}^{n} |x^k - \bar{x}^k|.$$

This completes the proof that Finsler manifolds have property I.

The main result of this section is the following.

THEOREM 1. *On a metric manifold of class C' with property I, a necessary and sufficient condition that an arc g be rectifiable is that every subarc covered by an admissible coördinate system (x) be given by functions $x^i = x^i(t)$ $a \leqq t \leqq b$ of bounded variation in $a \leqq t \leqq b$.*

To prove the necessity of the condition, let h be any subarc of g covered by an admissible coördinate system (x), and let $x^i = x^i(t)$ $a \leqq t \leqq b$ be the functions defining h. If g is rectifiable, so is h. From the first inequality of property I it follows easily by a summation process that $x^i(t)$ is of bounded variation in the interval $a \leqq t \leqq b$.

To prove the sufficiency of the condition, we first note that g can be covered by a finite number of admissible coördinate systems. Thus g can be divided into a finite number of arcs each in a single coördinate system. To prove the rectifiability of g, it is sufficient to prove the rectifiability of these subarcs. Let h be an arbitrary one of them, given by functions $x^i = x^i(t)$, by hypothesis of bounded variation. From the second inequality of property I it follows by a summation process that h is rectifiable.

COROLLARY. *On a Finsler manifold, a necessary and sufficient condition that an arc be rectifiable is that every subarc covered by an admissible coördinate system (x) be given by function $x^i = x^i(t)$ $a \leqq t \leqq b$ of bounded variation in the interval $a \leqq t \leqq b$.*

From this it follows that on a Finsler manifold integral arc length can be defined on any rectifiable curve, and every curve with integral arc length is rectifiable.

4. In euclidean space, the integral arc length of a rectifiable arc is equal to its metric arc length. This theorem is also true on Finsler manifolds.[2] We shall obtain a more general form of this theorem for Finsler manifolds. The generalization will lie in the fact that metric arc length, which is a limit of a sum of distances and hence is the limit of a sum of lengths of short extremal segments, can be computed as the sum of lengths of arcs belonging to a set of "paths."

---

[2] See a forthcoming paper by Marston Morse, *Functional topology and abstract variational theory*, Mémorial des Sciences Mathématiques (1938), Theorem 16.2.

Consider a family of curves (paths) in any n-dimensional metric manifold M of class C^3, with the following properties:[3]

(1) The family is to contain $2n - 2$ essential parameters. In a local coördinate system (x) the family is to be given by functions

$$x^i = h^i(s + \beta, \alpha_1, \cdots, \alpha_{2n-2})$$

where s is metric arc length along the curves, $(\alpha) = (\alpha_1, \cdots, \alpha_{2n-2})$ is a set of parameters determining the curve, and β is a parameter depending on the point from which s is to be measured. The functions $h^i(s + \beta, \alpha)$ and $h^i_s(s + \beta, \alpha)$ are to be of class $C^{(2)}$.

(2) One and only one curve of the family passes through any given point in any given direction. In a local coördinate system (x), the parameters (α) which give the curve passing through a given point (x) in a given direction (p) are to be functions of class $C^{(2)}$ of (x, p). This can be insured by assuming that the $(2n \times 2n - 1)$-matrix

$$\left\| \begin{matrix} h^i_s & h^i_{\alpha_\mu} \\ h^i_{ss} & h^i_{s\alpha_\mu} \end{matrix} \right\| \quad (i = 1, \cdots, n; \mu = 1, \cdots, 2n-2)$$

is of rank $2n - 1$.

A family of curves with these properties we shall call a *metric system of paths.*

As a consequence of the hypothesized properties, it can be shown that a metric system of paths has the following further properties.[4] The functions h^i can be put in the forms (1) and (2) of §2. Each point P of the manifold has a neighborhood simply covered by the paths issuing from P in which normal coördinates similar to those of (3) can be used. This neighborhood can always be taken to be of the form $s < \delta$, s being arc length measured from P along the paths issuing from P, and is then called the N-neighborhood of P. Each point has a neighborhood in which any pair of points can be joined by a unique path lying in that neighborhood, and in which there is no pair of conjugate points, so that the neighborhood is an N-neighborhood of each of its points. Hence for a compact set S there exists an ϵ such that if P, Q are any two points of S for which $\rho(P, Q) < \epsilon$, Q lies in the N-neighborhood of P.

The extremals of a Finsler manifold form an example of a metric system of paths. An interesting example of a metric system of paths not extremals consists of the set of circles of radius 1 in the euclidean plane, oriented say in the clockwise direction.

Now suppose we have any metric system of paths T on M. Let c be any arc of M, joining A to B. Consider any finite set of points $A = P_0, P_1, P_2, \cdots, P_m, P_{m+1} = B$ on c, ordered in the direction of increase of a parameter t increasing from A to B. Let Δt_k $(k = 0, \cdots, m)$ be the difference in the values of t at P_k and P_{k+1}. There exists an ϵ independent of k such that

[3] Cf. J. Douglas, *The general geometry of paths*, Annals of Math., vol. 29 (1927), p. 143.

[4] Cf. Douglas, loc. cit. and Whitehead, loc. cit.

if $\Delta t_k < \epsilon$, P_{k+1} lies in the N-neighborhood of P_k and a unique path of T passes from P_k to P_{k+1} in this neighborhood. Let l_k be the metric arc length of this path. Take max $\Delta t_k < \epsilon$, and form $\sum_0^m l_k$. If, no matter how the points of division P_k are taken, this sum approaches a unique limit L as max $\Delta t_k \to 0$, L is called the *path length* of c based on the paths T. Clearly if the set of paths T consists of geodesic arcs, the path length of c is metric arc length.

Our theorem is the following.

THEOREM 2. *On a Finsler manifold the path length L (based on any metric system of paths) of a rectifiable arc equals its integral arc length.*

This theorem will be proved by showing that the path length of a rectifiable arc c equals its metric arc length. For this purpose it is sufficient to prove the following lemma.

LEMMA. *On any compact subset S of a Finsler manifold, the ratio of the length $l(P, Q)$ of the path joining P to Q to the distance $\rho(P, Q)$ approaches the limit* 1 *as $\rho(P, Q) \to 0$, uniformly for all P, Q on S.*

If this were untrue, there would exist an ϵ and a sequence of pairs of points P_k, Q_k with $\rho(P_k, Q_k) \to 0$ and hence $l(P_k, Q_k) \to 0$, such that the ratio of $l(P_k, Q_k)$ to $\rho(P_k, Q_k)$ would differ from 1 by more than ϵ, for all k. Let P_0 be a limit point of P_k. We shall obtain a contradiction by showing that there exists a neighborhood σ of P_0 and a $\delta > 0$ such that any path of length $< \delta$ issuing from any point of σ makes the ratio in question lie within ϵ of 1.

Let (x) be a coördinate system about P_0, and let (y) be the geodesic normal coördinate system with a variable point $\overline{P}:(\bar{x})$ near P_0 as origin. The equations of paths can be written, in the coördinate system (x),

$$x^i = x^i(s, \bar{x}, v) \qquad (i = 1, \cdots, n)$$

where s is the length of the path, $(\bar{x})$ its initial point, and (v) a unit vector at $(\bar{x})$ representing the initial direction of the path, while the functions x^i are of class $C^{(2)}$. The transformation from (x) to (y) is given by

$$y^i = y^i(\bar{x}, x) \qquad (i = 1, \cdots, n)$$

and the equations of paths become

$$y^i = y^i[\bar{x}, x(s, \bar{x}, v)] = Y^i(s, x, v)$$

where the functions Y^i are of class C', and

$$Y^i(0, \bar{x}, v) \equiv 0$$
$$Y^i_s(0, \bar{x}, v) \equiv v^i.$$

The function $F(x, x')$ is transformed into a function $f(\bar{x}, y, y')$. The distance $\rho(\overline{P}, P)$ from $(\bar{x})$ to (x) is given as follows:

$$\rho(\overline{P}, P) = [\rho(\overline{P}, P)][f(\bar{x}, 0, p)] = f(\bar{x}, 0, p\rho) = f(\bar{x}, 0, y)$$

where (p) is the unit vector representing the initial direction of the geodesic from $\overline{P}$ to P. The ratio $\dfrac{l(\overline{P}, P)}{\rho(\overline{P}, P)}$ is

$$s/f(\bar{x}, 0, y) = s/f\left(\bar{x}, 0, \frac{y}{s}\right)\cdot s = 1/f\left(\bar{x}, 0, \frac{y}{s}\right).$$

Using the law of the mean, this becomes

$$1/f[\bar{x}, 0, Y_s(\theta s, \bar{x}, v)] \qquad \text{where } 0 \leqq \theta_i \leqq 1.$$

But since Y_s^i is continuous in s, uniformly for all unit vectors (v) and all $(\bar{x})$ on a compact neighborhood σ of P_0, and since $f(\bar{x}, 0, r)$ is continuous in (r), uniformly for all $(\bar{x})$ in σ, the uniform limit of the ratio as $s \to 0$ is

$$1/f(\bar{x}, 0, v) = 1.$$

This completes the proof.

For an interesting application of this theorem, we refer to the example already mentioned, that of the metric system of paths in the euclidean plane formed by the circles of radius 1 oriented in a clockwise direction. According to our theorem, one can obtain the length of an arc of class C' in the euclidean plane by taking the limit of the sum of inscribed circular arcs (less than semicircles) of radius 1 (instead of the usual straight line segments).

5. The theorem of the preceding section, combined with a result of J. Douglas,[5] leads to a further theorem. Douglas' result, when modified to hold for the non-analytic case in the large, states essentially that if a metric system of paths T exists on a metric manifold M of class $C^{(4)}$, a unique Finsler integral can be exhibited such that M becomes a Finsler manifold in the sense of §2, and the arc length furnished by this integral coincides with the given metric arc length along the paths of T. Of course the given system of paths T is in general not the system of extremals of the Finsler manifold. However, the path length of any rectifiable arc as computed by means of T will be the same as the arc length arising from the integral, by Theorem 2.

If M is a metric manifold with T as its geodesic system, we have the following theorem.

THEOREM 3. *If the geodesic arcs of a metric manifold M of class $C^{(4)}$ form a metric system of paths, then M is a Finsler manifold. More precisely, a Finsler metric can be exhibited such that arc length of rectifiable arcs as calculated by it is the same as the original metric arc length on M, and distances between nearby points are the same in both metrics.*

If M is a complete, convex metric manifold, then all distances will be the same in both metrics. For on such a space every pair of points can be joined by a geodesic arc, so that distance between two points is always equal to the greatest lower bound of length of arcs joining the two points.

UNIVERSITY OF MICHIGAN.

[5] Loc. cit., pp. 150–152.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

THE LINEAR HOMOGENEOUS GROUP[1]

BY JOEL BRENNER

(Received February 3, 1937)

Introduction.

1. Fundamental generating theorems ($p \geqq 2$).
2. Canonical form for $n = 2$, $s > 0$, $p > 2$.
3. Fundamental theorem for $n = 2$, $s = 0$, $p > 2$.
4. Fundamental theorem for $n \geqq 2$, $s > 0$, $p > 2$.
5. Special normal subgroups for $n = 2$, $p = 2$.
6. Fundamental theorem for $n = 2$, $s > 0$, $p = 2$.
7. Fundamental theorem for $n > 2$, $s > 0$, $p = 2$.
8. Fundamental theorem for $n = 2$, $s = 0$, $p = 2$.
9. Fundamental theorem for $n > 2$, $s = 0$, $p \geqq 2$.
10. The normal subgroups when $p > 2$. Summary.
11. The normal subgroups when $n = 2$, $p = 2$.
12. The normal subgroups when $n > 2$, $p = 2$.
13. Relations among A^p; miscellaneous results ($p > 2$).

Introduction. The object of this article is to describe the normal subgroups of the group $\mathfrak{G}_{p,n,r}$ of non-singular n^2 matrices whose elements are integers mod p^r, $n \geqq 2$, p a prime. The group $\mathfrak{G}$ is the group of automorphisms of the Abelian group of order p^{nr} and type $(r, r, \cdots r)$.[2]

For each matrix $A = (a_{ij})$ of $\mathfrak{G}$ we define[3] $r_{ij}(A) = \log_p (a_{ij}, p^r)$ = exponent of highest power ($\leqq p^r$) of p in a_{ij}; $q_{ij}(A) = \log_p (a_{ii} - a_{jj}, p^r)$; $s(A) = \min_{i \neq j}(r_{ij}(A), q_{ij}(A))$.[3a] We denote by $\mathfrak{S}_A$ the (normal) group generated by A and all its conjugates under $\mathfrak{G}$. We are able to prove (2–4, 7, 9) that if $np \geqq 6$, $\mathfrak{S}_A$ contains a certain normal (in $\mathfrak{G}$) subgroup $\mathfrak{N}_{s(A)}$ which consists of all matrices $X = (x_{ij})$ of $\mathfrak{G}$ satisfying $\det X = 1$ (p^r), $s(X) \geqq s(A)$, $x_{ii} = 1$ (p^r) $(i = 1, \cdots, n)$. The group $\mathfrak{S}_A$ is moreover contained in the (normal) subgroup $\mathfrak{N}^*_{s(A)}$ consisting of all matrices X with $s(X) \geqq s(A)$; the normal groups $\mathfrak{N}$ thus satisfy $\mathfrak{N}^*_s \supseteq \mathfrak{N} \supseteq \mathfrak{N}_s$ for appropriate s, and description of $\mathfrak{N}$ follows from this fact (10, 12). Finally occur some miscellaneous results which may have a bearing on the characteristic subgroups of $\mathfrak{G}$ (14).

---

[1] Revision and amplification of a doctoral thesis written at Harvard University. I am indebted to Mr. Garrett Birkhoff and Prof. M. H. Stone for valuable advice and encouragement during this study.

[2] W. Burnside, Theory of Groups of Finite Order, Cambridge University Press, 1897, p. 242.

[3] Highest common factor of a and b is (a, b).

[3a] We have $s(AB) = \min s(A), s(B)$ unless $s(A) = s(B)$; $s(AB) \geqq s(A)$; $s(CAC^{-1}) = s(A)$.

1. **Fundamental generating theorems** ($p \geqq 2$). One of the theorems in this section was stated in the case $r = 1$ by C. Jordan.[4]

Subgroups of $\mathfrak{G}$ are denoted throughout by German capital letters; matrices of $\mathfrak{G}$ by Roman capitals. In particular W_{ij} $(i \neq j)$ is the matrix which differs from the unit matrix $\mathbf{1}$ only in the (i, j)th place, where 1 stands (instead of 0); W_{ij}^m $(i \neq j)$ is the mth power of W_{ij}, and differs from $\mathbf{1}$ only in the (i, j)th place, where m stands (instead of 0); $V_i^{(\alpha)}$ differs from $\mathbf{1}$ only in the (i, i)th place, where α $((\alpha, p) = 1)$ stands (instead of 1). The inverse of $V_i^{(\alpha)}$ is $V_i^{(1/\alpha)}$, where $1/\alpha$ is the integer satisfying $\alpha(1/\alpha) = 1\ (p^r)$. An s-matrix is a matrix A with $s(A) = s$ $(0 \leqq s \leqq r)$.

In writing a matrix A as a square array, a_{ij} is written in the ith row and the jth column. In the following, it will be convenient to know that $W_{ij}^m A W_{ij}^{-m}$ does not differ from A except possibly in the ith row or in the jth column; and $V_i^{(\alpha)} A V_i^{(1/\alpha)}$ does not differ from A in the main diagonal, nor elsewhere except possibly in the ith row or in the ith column.

THEOREM 1.1. *If $n = 2$ and p is any prime $\geqq 2$, then W_{21}^w and W_{12}^w $(w = p^s)$ together generate all matrices (p_{ij}) such that (1) $|p_{ij}| = 1\ (p^r)$, $p_{ij} = 0\ (p^s)$ when $i \neq j$; $p_{ij} = 1\ (p^{2s})$ when $i = j$. These matrices form a group $\mathfrak{C}_s$.*

PROOF. Set $F = W_{12}^{mw} W_{21}^{lw} W_{12}^{m'w}$. By choice of m, m', and l, with $(l, p) = 1$, we can make this an arbitrary matrix (p_{ij}) subject to the restriction (1) and the added restriction $(p_{21}, p^r) = p^s$. If $p > 2$, or if $s = r$, these matrices form a majority of the matrices of $\mathfrak{C}_s$, so that the whole group $\mathfrak{C}_s$ must be generated. If $p = 2$ and $s < r$, these matrices form half the group $\mathfrak{C}_s$; to get the necessary majority we need merely observe that $\mathbf{1}$ is also generated. Hence, theorem 1.1 is true.

THEOREM 1.2. *The set $\{W_{ij}^w\}$ $(w = p^s)$ $(i, j = 1, \cdots, n, i \neq j)$ generates all matrices (p_{ij}) such that (1) holds. These matrices form a group $\mathfrak{C}_s$.*

PROOF. It is obvious that $\mathfrak{C}_s$ is a group. Let (p_{ij}) be any matrix satisfying (1) and with $p_{nn} = 1(p^r)$. We first show that such a matrix can be generated by the set. According to the induction hypothesis, it is possible to generate[5] diag $(Q, \mathbf{1})$ where Q is the $(n-1)^2$ box (q_{ij}) defined by $q_{ij} = p_{ij} - p_{nj}p_{in}$ $(i, j = 1, \cdots, n-1)$, because $|q_{ij}| = 1\ (p^r)$; $q_{ij} = 0\ (p^s)$ when $i \neq j$, $q_{ij} = 1\ (p^{2s})$ when $i = j$. Now consider the product

$$(a_{ij}) = \prod_{k=1}^{n-1} W_{nk}^{p_{nk}} \operatorname{diag}(Q, \mathbf{1}) \prod_{l=1}^{n-1} W_{ln}^{p_{ln}}.$$

It can be verified that $a_{nn} = 1$; $a_{ij} = p_{ij}$ $(i, j \neq n, n)$. Hence our set of generators generates every matrix satisfying (1) and with $p_{nn} = 1(p^r)$. Now let an arbitrary matrix $P' = (p'_{ij})$ satisfying (1) be given. Since $p'_{nn} = 1(p^{2s})$, hence also $p'^{-1}_{nn} = 1(p^{2s})$. From what has already been proven, the set will

[4] Jordan, Traité des Substitutions, 1870, p. 143 and thereabouts.

[5] To save space, the above abbreviation is used for a matrix such that the numbers (or boxes) along the main diagonal are those written, and all the other elements of the matrix are 0.

generate the matrix[5] diag $(1, 1, \cdots, 1, p'^{-1}_{nn}, p'_{nn}, 1)$ hence also the matrix diag $(1, 1, \cdots, 1, p'_{nn}, p'^{-1}_{nn})$, by symmetry. The product of the latter by P' is a matrix P which has $p_{nn} = 1(p^r)$, and satisfies (1). Hence P can be generated by the set, and the same must be true of P', Q.E.D.

THEOREM 1.3. *A subgroup of $\mathfrak{G}$ is normal if it is fixed under all W_{ij} and $V_i^{(\alpha)}$.* For W_{ij} $(i, j = 1, \cdots, n)$ generate all matrices of determinant $= 1(p^r)$; any matrix of $\mathfrak{G}$ of determinant Δ is the product of $V_1^{(\Delta)}$ by a suitable matrix of determinant 1.

COROLLARY. *$\mathfrak{C}_0$ is normal.*

THEOREM 1.4. *The commutator subgroup of $\mathfrak{G}$ is $\mathfrak{C}_0$ if $p > 2$.*

PROOF: The commutator subgroup of $\mathfrak{G}$ obviously is contained in $\mathfrak{C}_0$. Let α be any number such that $\alpha \not\equiv 0$, $\alpha \not\equiv 1(p)$. Then $[W_{ij}V_j^{(\alpha)}W_{ij}^{-1}V_j^{(1/\alpha)}]^{1/(1-\alpha)} = W_{ij}$. (The exponent is an integer.) Hence the commutator subgroup contains $\mathfrak{C}_0$.

THEOREM 1.4a. *If $n \geqq 3$, $p = 2$, the commutator subgroup of $\mathfrak{G}$ in $\mathfrak{C}_0$.*

PROOF. $W_{13}W_{21}W_{13}^{-1}W_{21}^{-1} = W_{23}$. (See lemma 5.6.)

THEOREM 1.5. *The normal subgroup generated by $\mathfrak{C}_s$ (i.e. the cross-cut of all the normal subgroups of $\mathfrak{G}$ which contain $\mathfrak{C}_s$) is $\mathfrak{N}_s$.*[5a]

PROOF: It is easy to verify that $\mathfrak{N}_s$ is a normal subgroup by noting that it is invariant under W_{ij} and $V_i^{(\alpha)}$.

Reading down the main diagonal of the transform T of $\prod_1^{n-1} W_{i,i+1}^{w}$ $(\epsilon\mathfrak{C}_s)$ $(w = p^s)$ by

$$W_{n,n-1}^{t_{n-1}}W_{n-1,n-2}^{t_{n-2}}\cdots W_{21}^{t_1} \qquad (t_j = \textstyle\sum_1^j m_i),$$

where m_i are integers, we find

$$1 + m_1p^s,\ 1 + m_2p^s,\ \cdots,\ 1 + m_{n-1}p^s,\ 1 - (\textstyle\sum_1^{n-1} m_i)p^s. \tag{4}$$

Now let any matrix $A = (a_{ij})$ of $\mathfrak{N}_s$ be given. We have $\det A = 1\ (p^r)$. Each of the $n!$ terms in the expansion of $\det A$ except $a_{11}a_{22}\cdots a_{nn}$ is $= 0\ (p^{2s})$. Hence $a_{11}a_{22}\cdots a_{nn} = 1\ (p^{2s})$ and since $a_{ii} = 1(p^s)$, we have $\sum_{i=1}^{n}(a_{ii} - 1) = 0\ (p^{2s})$. Hence, if m_i $(i = 1, \cdots, n-1)$ are properly chosen, the elements of the main diagonal of A are congruent to the series (4), modulo p^{2s}. The matrix AT^{-1} belongs to $\mathfrak{C}_s$, whence A can be generated by the conjugates of the elements of $\mathfrak{C}_s$.

2. Canonical form for $n = 2$, $s > 0$, $p > 2$.

LEMMA 2.1. *Let $n = 2$, $s = s(A) > 0$, $p > 2$. Then $A = (a_{ij})$ has a conjugate $B = (b_{ij})$ with* (1) $b_{21} = p^s$, $b_{11} = b_{22}\ (p^r)$. *A necessary and sufficient condition that two s-matrices satisfying* (1) *be conjugate is that they be identical.*

PROOF. EXISTENCE. Assume without loss of generality that $r_{21}(A) \leqq r_{12}(A)$. (Otherwise transform by (0, 1, 1, 0).[6]) With $r_{21}(A) \leqq r_{12}(A)$, assume without loss of generality that $r_{21}(A) = s$. (Otherwise we have $q_{12}(A) < r_{21}(A) \leqq r_{12}(A)$

[5a] Defined in introduction.

[6] To save space, the 2^2 matrix (a_{ij}) is printed $(a_{11}, a_{21}, a_{12}, a_{22})$.

and we can transform by W_{21}.) With $r_{21}(A) = s$, assume without loss of generality that $a_{21} = p^s$ (p^r). (Otherwise transform by $V_1^{(\alpha)}$ with appropriate α.) With $a_{21} = p^s$ (p^r), B is $W_{12}^k A W_{12}^{-k}$, where k satisfies $2kp^s = a_{11} - a_{22}$ (p^r). Thus every s-matrix has a conjugate satisfying (1) if $n = 2$, $p > 2$, $s > 0$.

Uniqueness proof omitted.

3. **Fundamental theorem for $n = 2$, $s = 0$, $p > 2$.**

LEMMA 3.1. *(Canonical form for $n = 2$, $s = 0$, $p \geqq 2$.) Let $n = 2$, $s = s(A) = 0$, $p \geqq 2$. Then $A = (a_{ij})$ has a conjugate $B = (b_{ij})$ with* (1) $b_{21} = 1$, $b_{11} = 0$ (p^r). *A necessary and sufficient condition that two 0-matrices satisfying* (1) *be conjugate is that they be identical.*

PROOF: Existence. As in lemma 2.1, assume without loss of generality that $a_{21} = p^s = 1$ (p^r). Then B is $W_{12}^m A W_{12}^{-m}$, $m = a_{11}$.

Uniqueness. $b_{12} = -\det B$ (p^r). For the invariance of b_{22}, a direct proof may be used.

LEMMA 3.2. *If $s(A) = 0$, $n = 2$, $p > 2$, then $\mathfrak{S}_A$ contains a matrix B with $r_{21}(B) = r$, $r_{12}(B) = 0$.*

PROOF: Assume that A is in canonical form. If $(a_{22}, p) = 1$ then $B = A \cdot W_{12}^{-m} A W_{12}^m$ $(m = a_{22})$. If $(a_{22}, p) = p$, $p > 3$, then

$$B = W_{12}^{-1} A W_{12} \cdot W_{12}^{-m} V_1^{(2)} A V_1^{(1/2)} W_{12}^m \qquad (m = 1 + a_{22}/2).$$

If $(a_{22}, p) = p$, $p = 3$, we have two steps. First we prove $\mathfrak{S}_A \supset \mathfrak{N}_1$. Set $m = a_{22} - 3$. Then $a_{22} + m$ is divisible by 3. Let $C = W_{12}^m A W_{12}^{-m} \cdot A$. Since $s(C) = 1$, hence $\mathfrak{S}_A \supset \mathfrak{N}_1$ by theorem 2.2. Now consider

$$E = V_1^{(\alpha)} W_{12}^m A W_{12}^{-m} W_{12}^{-1} D W_{12} V_1^{(1/\alpha)},$$

where D is obtained by permuting[7] 1, 2 in the inverse of A; $\alpha = -a_{12}$, $m = 1/a_{12} - a_{22}/a_{12} - 1$. We have $E = (0, 1, -1, e_{22})$,[6] and hence E is in $\mathfrak{N}_0$ but not in $\mathfrak{N}_1$. Now $\mathfrak{S}_E$ certainly is $\mathfrak{N}_0$ if $e_{22} \not\equiv 0$ (3) (see above). If $e_{22} = 0$ (3), then E is of order 4 mod $\mathfrak{N}_1$. Since (order $\mathfrak{N}_0$)/(order $\mathfrak{N}_1$) = 12, the fact that the four cosets $\mathfrak{N}_1$, $E\mathfrak{N}_1$, $E^2\mathfrak{N}_1$, $E^3\mathfrak{N}_1$ do not form a normal subgroup proves that $\mathfrak{S}_E = \mathfrak{N}_0$ in this case too. (In fact $W_{12} E W_{12}^{-1}$ is not in any of these four cosets.)

LEMMA 3.3. *If $s = s(B) = 0$, $r_{21}(B) = r$, $r_{12}(B) = 0$, $p > 2$, then $\mathfrak{S}_B \supseteq \mathfrak{N}_s$.*

PROOF: $V_2^{(2)} B V_2^{(1/2)} B^{-1} = W_{12}^m$, where $(m, p) = 1$.

THEOREM 3.4. *When $p > 2$, $s(A) = 0$, $n = 2$, then $\mathfrak{S}_A \supseteq \mathfrak{N}_{s(A)}$.*

4. **Fundamental theorem for $n \geqq 2$, $s > 0$, $p > 2$.**

DEFINITION 4.1. *The matrix $A = (a_{ij})$ is said to be in pseudo-normal form if* (a) $a_{ij} = 0$ (p^r) *whenever $i > j + 1$, and* (b) $a_{21} = p^{s(A)}$ (p^r). *The matrix $A = (a_{ij})$ is in normal form if $a_{ij} = 0$ (p^r) whenever $i > j$.*

DEFINITION 4.2. *The matrix A is an open product (W. A. Manning) if for*

[7] That is, transform by a matrix of determinant ± 1, with n units and $n^2 - n$ zeros for elements.

some k, $0 < k < n$, (a) $a_{ij} = 0\ (p^r)$ $(i > k, j \leqq k)$; (b) $a_{ij} = 0\ (p^r)$ $(i \leqq k, j > k)$; *in pictures* $A = (A_1, \mathbf{0}, \mathbf{0}, A_2)$.[6] *The matrix* A *is a semi-open product if condition* (b) *is neglected; in pictures* $A = (A_1, \mathbf{0}, T, A_2)$.

LEMMA 4.3. *If* A *and* B *are open (semi-open) products for the same* k, *then so is* AB.

LEMMA 4.4. *Every matrix has a conjugate in pseudo-normal form* $(p \geqq 2)$.

REMARK: The transformation is usually possible in many ways. When it is desired in subsequent proofs to "transform B into p.n. form," it is understood that the algorithm of this proof should be used.

PROOF: Consider $B = (b_{ij})$. We have

$$s = s(B) = \min_{i \neq j} \log_p (b_{ij}, b_{ii} - b_{jj}, p^r).^8$$

If it is not true that $p^s = (b_{21}, b_{22} - b_{11}, p^r)$, then we first transform by a permutation matrix[7] to bring this about. Then as in lemma 2.1 we reach by transformations a matrix $C = (c_{ij})$ with $c_{21} = p^s$. Now $r_{21}(C) = s = \min_{\theta>1} r_{\theta 1}(C)$. Transform C by $T_1 = W_{32}^{m_3} \cdots W_{n2}^{m_n}$, where m_i are so chosen that $m_i p^s = c_{i1}(p^r)$; the resulting matrix $T_1 C T_1^{-1} = D = (d_{ij})$ satisfies $d_{21} = p^s$; $d_{31} = d_{41} = \cdots = d_{n1} = 0\ (p^r)$. Suppose $r_{k2}(D) = \min_{\theta>2} r_{\theta 2}(D)$. If we permute[7] the subscripts k and 3, getting $E = (e_{ij})$ (if $k = 3$, $E = D$), we can transform E by $T_2 = W_{43}^{m_4'} \cdots W_{n3}^{m_n'}$, where m_i' are so chosen that $m_i' e_{32} = e_{i2}\ (p^r)$ (all soluble as $r_{i2}(E) \geqq r_{32}(E)$). Suppose $r_{l3}(T_2 E T_2^{-1}) = \min_{\theta>3} r_{\theta 3}(T_2 E T_2^{-1})$. Permute the subscripts l and 4, etc.; repeating this process we reach pseudo-normal form eventually.

LEMMA 4.5. *If* A *is in pseudo-normal form, and* $s = s(A) > 0$, $p > 2$, $n > 2$, *then* $\mathfrak{S}_A$ *contains a matrix* B *in pseudo-normal form with* $r_{32}(B) = r$.

PROOF: Set $\alpha = 2$, $m = 1$; solve $(1/\alpha)a_{22} + \beta a_{33} = 0\ (p^r)$ for β. If $A' = V_3^{(1/\beta)} A V_3^{(\beta)} W_{12}^{-m} V_2^{(1/\alpha)} A V_2^{(\alpha)} W_{12}^{m}$ be transformed into p.n. form the resulting matrix will be of the same type as A, only with r_{32} increased. Continuing in this way, we reach B eventually.

LEMMA 4.6. *Let* $B = (b_{ij})$ *be as defined in the previous lemma* $(p > 2, n \geqq 2)$. *Then* $\mathfrak{S}_B$ *contains a matrix* C *in normal form, with* $r_{12}(C) = s(B)$, *and with the lower right* $(n-2)^2$ *box of* C *the* $(n-2)^2$ *unit matrix.*

PROOF: By lemma 2.1, we may assume $b_{11} = b_{22}\ (p^{s+1})$. C is

$$B^{-1} W_{12}^{m} V_2^{(1/\alpha)} B V_2^{(\alpha)} W_{12}^{-m},$$

where $\alpha = p^s + 1$, $m = (1/\alpha)b_{22}\ (p^r)$.

LEMMA 4.7. *Let* $C = (c_{ij})$ *be as defined in the previous lemma. Then* $\mathfrak{S}_C$ *contains* W_{12}^{m}, $(m, p^r) = p^s$.

PROOF: Assume without loss of generality that $c_{13} = c_{14} = \cdots = 0$. (Otherwise transform by $W_{23}^{m_3} W_{24}^{m_4} \cdots W_{2n}^{m_n}$.) Then $W_{12}^{m} = V_2^{(2)} C V_2^{(1/2)} C^{-1}$.

THEOREM 4.8. *If* $s(A) > 0$, $p > 2$, $n \geqq 2$, *then* $\mathfrak{S}_A \supseteq \mathfrak{N}_{s(A)}$.

5. Special normal subgroups for $n = 2$, $p = 2$. As the theorems of this paragraph are subject to direct verification, their proof is merely sketched.

LEMMA 5.1. *The least normal subgroup $\mathfrak{J}_s$ of $\mathfrak{G}$ containing both $\mathfrak{N}_{s+1}$ and $A_s = (1 + 2^s, 2^s, 0, 1/(1 + 2^s))$ $(0 < s < r)$[6] is of index 2 under $\mathfrak{N}_s$ and contains $\mathfrak{N}_{s+1}$ with index 4.*

PROOF: Representatives of the four cosets of $\mathfrak{J}_s/\mathfrak{N}_{s+1}$ are **1**, A_s $(1 + 2^s, 0, 2^s, 1/(1 + 2^s))$, $(1 + 2^{2s}, 2^s, 2^s, 1)$. We can show that $\mathfrak{J}_s$ is normal by showing that one coset is transformed into another by W_{21} or W_{12}. $V_1^{(\alpha)}$ fixes each coset. $\mathfrak{J}_s/\mathfrak{N}_{s+1}$ is the four group; $\mathfrak{N}_s/\mathfrak{N}_{s+1}$ is of type $(1, 1, 1)$ $(0 < s < r)$.

LEMMA 5.2. *$\mathfrak{N}_{s+1}$ and its coset represented by $A = (1 \pm 2^s, 0, 0, 1/(1 \pm 2^s))$ fill out a normal subgroup $\mathfrak{H}_{s+1}$. There is no normal subgroup between $\mathfrak{N}_s$ and $\mathfrak{H}_{s+1}$ $(0 < s < r)$.*

LEMMA 5.3. *If $1 < s < r - 1$, the least normal subgroup $\mathfrak{N}$ containing both $\mathfrak{J}_{s+1}$ and A of the previous lemma is $\mathfrak{H}_{s+1}$.*

PROOF: $\mathfrak{H}_{s+1}/\mathfrak{J}_{s+1}$ is of order four, and $\mathfrak{H}_{s+1} \supseteq \mathfrak{N} \supseteq \mathfrak{J}_{s+1}$. Also, $\mathfrak{N} \supseteq \mathfrak{N}_{s+1}$. Since $\mathfrak{N}$ could not be $\mathfrak{N}_{s+1}$, it must be $\mathfrak{H}_{s+1}$.

LEMMA 5.3a. *The groups $\mathfrak{L}_2 = \{\mathfrak{J}_2, -\mathbf{1}\}$ and $\mathfrak{P}_2 = \{\mathfrak{J}_2, \text{diag } (3, \frac{1}{3})\}$ are normal $(r > 2)$.*

LEMMA 5.4. *The least normal subgroup $\mathfrak{K}$ containing $\mathfrak{J}_1$ and $(0, 1, -1, -1)$ contains $\mathfrak{J}_1$ with index 3 and is contained in $\mathfrak{N}_0$ with index 4.* The three representatives of $\mathfrak{K}/\mathfrak{J}_1$ are **1**, $(0, 1, -1, -1)$, $(-1, -1, 1, 0)$.

LEMMA 5.5. *The subgroup $\mathfrak{F}$ generated by $\mathfrak{K}$ and $(0, -1, 1, 1)$ is of index 2 under $\mathfrak{N}_0$ and is also normal; $\mathfrak{F}$ contains $\mathfrak{N}_1$ (with index 3).*

PROOF: Since $(-1, -2, 0, -1)$ is in $\mathfrak{J}_1$, we see that the product $(-1, -2, 0, -1)$ $(0, -1, 1, 1)^3 = W_{21}^2$ is in $\mathfrak{F}$. Hence $\mathfrak{F}$ contains $\mathfrak{N}_1$.

LEMMA 5.6. *The commutator subgroup $\mathfrak{C}$ for $n = 2$ contains $\mathfrak{F}$ $(p = 2)$.*

PROOF: We have $V_1^{(1/3)} W_{12} V_1^{(3)} W_{12}^{-1} = W_{12}^2$, so that by theorem 1.5, $\mathfrak{C} \supseteq \mathfrak{N}_1$. Further, the product of the commutator $(1, 0, -1/3, 1)(1, -3, 0, 1)$ $(1, 0, 1/3, 1)$ $(1, 3, 0, 1) = (0, -3, 1/3, 3)$ by the matrix $(3, 0, 0, 1/3)$ from $\mathfrak{N}_1$ is $(0, -1, 1, 1)$. Hence $\mathfrak{C} \supseteq \mathfrak{F}$. [As a matter of fact, $\mathfrak{C} = \mathfrak{F}$ when $n = 2$, $p = 2$.]

6. Fundamental theorem for $n = 2$, $s > 0$, $p = 2$.

LEMMA 6.1. *(Canonical form for $n = 2$, $s > 0$, $p = 2$.) If $n = 2$, $p = 2$, $0 < s = s(B) < r$, then B has a conjugate $A = (a_{11}, 2^s, a_{12}, a_{11})$,[6] $0 < a_{11} < 2^{r-1}$, or a conjugate $A' = (a'_{11}, 2^s, a'_{12}, a'_{11} + 2^s)$, $0 < a'_{11} < 2^{r-1}$; A is not conjugate to A'; and a necessary and sufficient condition that two matrices of the above form be conjugate is that they be identical.*

As in lemma 2.1, B has a conjugate $C = (c_{ij})$ with $c_{21} = 2^s$. If $c_{11} = c_{22}$ (2^{s+1}), then C has a conjugate A. For if k is so chosen that $c_{11} - c_{22} = 2^{s+1}k$ (2^r), then one of W_{12}^k and $W_{12}^{k'}$ $(k' = k + 2^{r-s-1})$ transforms C as desired. If $c_{11} \neq c_{22}$ (2^{s+1}), then C has a conjugate A'. For if k is so chosen that $c_{11} - c_{22} = 2^{s+1}k - 2^s$ (2^r), then one of W_{12}^k and $W_{12}^{k'}$ $(k' = k + 2^{r-s-1})$ transforms C as desired.

UNIQUENESS: That A is not conjugate to A' is obvious from the fact that any conjugate B of A has $q_{12}(B) = s + 1$, while any conjugate B' of A' has $q_{12}(B') = s$. (See theorem 1.3.) Remainder of uniqueness proof omitted.

LEMMA 6.2. *If* $A = (a_{11}, 2^s, a_{12}, a_{22})$ *and* $s = s(A) > 0$, $q_{12}(A) > s$, $p = 2$, *then the least normal subgroup* $\mathfrak{S}_A$ *containing* A *contains also a matrix* $F = (f_{11}, 2^s, 0, f_{22})$, *with* $q_{12}(F) > s$.

PROOF: Form $D = (d_{ij}) = V_1^{(\alpha)} A V_1^{(1/\alpha)} W_{12}^m A W_{12}^{-m}$; where $a_{11} + \alpha a_{22} = 2^s (2^r)$, $m = -(1/\alpha)$. Transformation of D by $(0, \beta, 1, 0)$ with appropriate β yields F as required.

LEMMA 6.3. *The least normal subgroup* $\mathfrak{S}_{A'}$ *containing* $A' = (a'_{11}, 2^s, a'_{12}, a'_{22})$ $(s = s(A') > 0, p = 2)$, $a'_{11} \not\equiv a'_{22}$ (2^{s+1}) *contains also a matrix* $F = (f_{11}, 2^{s+1}, 0, f_{22})$ *with* $f_{11} = f_{22}$ (2^{s+1}).

PROOF: If we form the product D' corresponding to D in lemma 6.2, and solve $a_{11} + \alpha a_{22} = 2^{s+1}$ (2^r) for α, and set $m = -2(1/\alpha)$, the same subsequent step leads to a matrix F as required.

LEMMA 6.4. *The least normal subgroup* $\mathfrak{S}_F$ *containing* $F = (f_{11}, 2^s, 0, f_{22})$ $(0 < s = s(F) < r, p = 2)$ *contains* W_{21}^{s+1}, *and hence* $\mathfrak{N}_{s+1}$.

PROOF: Set $\alpha = 2f_{11} + 1$. Then $F^{-1} V_1^{(\alpha)} F V_1^{(1/\alpha)} = W_{21}^{s+1}$.

LEMMA 6.5. *The least normal subgroup* $\mathfrak{S}_F$ *containing* $F = (f_{11}, 2^s, 0, f_{22})$ $(s = s(F) > 0, p = 2)$ *contains also a matrix* $G = (g_{11}, 2^s, 0, g_{22})$ *with* $q_{12}(G) > s$.

PROOF: If $B = F(0, 1, 1, 0) F(0, 1, 1, 0)^{-1}$ be reduced to canonical form, it will be found that B satisfies $s(B) = s$ and that $q_{12}(B) > s$; hence lemma 6.2 is applicable.

NOTE: The same conclusion may be drawn if merely $f_{12} = 0$ (2^{s+1}) instead of (2^r) as here (same proof). Thus unless $r_{21}(A') = r_{12}(A') = q_{12}(A')$, $\mathfrak{S}_{A'}$ contains G above $(s = s(A') > 0)$.

LEMMA 6.6. *The least normal subgroup* $\mathfrak{S}_A$ *containing* $A = (a_{11}, 2^s, 0, a_{22})$ $(s = s(A) > 0, p = 2)$ *contains also* $(1 + 2^s, 2^s, 0, 1/(1 + 2^s))$.

PROOF: By lemma 6.5, assume without loss of generality that $a_{11} = a_{22}$ (2^{s+1}). By multiplying two conjugates of A as in Lemma 6.2, we reach another matrix $B = (b_{11}, 2^s, 0, b_{22})$ where $b_{11} = a_{11}^2$, $b_{22} = a_{22}^2$ (2^s), as is easily verified. Iteration yields a matrix $C = (c_{11}, 2^s, 0, c_{22})$ where $c_{11} = a_{11}^4$, $c_{22} = a_{22}^4$ (2^s). Repeated iteration shows that $\mathfrak{S}_A$ contains a matrix $D = (d_{11}, 2^s, 0, d_{22})$ with $d_{11} = d_{22} = 1$ (2^s). Then further iteration will give a matrix of the same type, only with determinant $= 1$ (2^r): $D' = (1 + k2^s, 2^s, 0, 1/(1 + k2^s))$. If k is odd, then we note that $E = ((1 + 2^s)/(1 + k2^s), 0, 0, (1 + k2^s)/(1 + 2^s))$ is in $\mathfrak{N}_{s+1}$, so that $D'E$ is in $\mathfrak{S}_A$, Q.E.D. If k is even, then pick $E = (1 + k2^s, 0, 0, 1/(1 + k2^s))$, and verify $D'E = W_{12}^w$ $(w = 2^s)$, so that $\mathfrak{S}_A$ contains $\mathfrak{N}_s$.

From lemmas 6.1–6.6, we conclude

THEOREM 6.7. *When* $n = 2$, $s = s(A) > 0$, *then* (a) $\mathfrak{S}_A$ *contains* $\mathfrak{I}_{s+1}$ $(s < r - 1)$; (b) *unless* $r_{21}(A) = r_{12}(A) = q_{12}(A)$, $\mathfrak{S}_A$ *contains* $\mathfrak{I}_s$ $(s < r)$.

LEMMA 6.8. *When* $n = 2$, $s = s(A) > 0$, *then if* $r_{21}(A) = r_{12}(A) = q_{12}(A)$, $\mathfrak{S}_A$ *contains one of* $(1 \pm 2^s, 0, 0, 1/(1 \pm 2^s))$.

PROOF: Set $W_{21} A W_{21}^{-1} A^{-1} = B = (b_{ij})$. We find that $b_{12} = 0$ or 2^{s+1}; $b_{21} = 0$ or 2^{s+1}; $b_{11} = 1 \pm 2^s$ (2^{s+2}). Therefore mod $\mathfrak{N}_{s+2}$ B has one of the six forms obtained by changing (2^{s+2}) to (2^r) in the above. (Note 1: if $b_{ij} = b'_{ij}$ (2^{s+2})

$(i, j = 1, \cdots, n)$, then $(b_{ij})(b'_{ij})^{-1}$ is in $\mathfrak{N}_{s+2}$.) (Note 2: if b_{12}, b_{21}, b_{11} are known (2^r), b_{22} is uniquely determined from $\det B = 1(2^r)$.) In the case of each of these six forms it is possible to show by direct computation that B and one of $(1 \pm 2^s, 0, 0, 1/(1 \pm 2^s))$ lie in the same coset of $\mathfrak{I}_{s+1}$. (Note 3: B and C lie in the same coset of $\mathfrak{I}_{s+1}$ if BC^{-1} is in $\mathfrak{I}_{s+1}$. A matrix D is in $\mathfrak{I}_{s+1}$ if and only if the product of D by one of the matrices corresponding to the ones in lemma 5.1 lies in $\mathfrak{N}_{s+2}$.)

THEOREM 6.9. *When $n = 2$, $s = s(A) > 1$, then $\mathfrak{S}_A \supset \mathfrak{H}_{s+1}$ or else $\mathfrak{S}_A \supseteq \mathfrak{I}_s$. When $n = 2$, $s(A) = 1$, then $\mathfrak{S}_A \supset \mathfrak{L}_2$ or else $\mathfrak{S}_A \supset \mathfrak{P}_2$ or else $\mathfrak{S}_A \supseteq \mathfrak{I}_1$.*

7. **Fundamental theorem for $n > 2$, $s > 0$, $p = 2$.** In the following chain of four lemmas, we show that if A is in p.n. form and $q_{12}(A) = s(A) > 0$, then $\mathfrak{S}_A$ contains a matrix B in p.n. form with $s(B) = s(A)$, $q_{12}(B) > s(A)$.

LEMMA 7.1. *If $A = (a_{ij})$ is in pseudo-normal form, and $q_{12}(A) = s(A) > 0$, $r_{32}(A) = s$, then $\mathfrak{S}_A$ contains a matrix B in pseudo-normal form with $q_{12}(B) > s(A)$, $s(B) = s(A)$ $(n > 2, s = s(A) > 0, p = 2)$.* Proof: $W_{23}AW_{23}^{-1}$ is such a matrix.

LEMMA 7.2. *If $A = (a_{ij})$ is in pseudo-normal form, and $q_{12}(A) = s(A) = s > 0$; $r_{32}(A) > s$; $r_{12}(A) > s$, then $\mathfrak{S}_A$ contains a matrix B in pseudo-normal form with $s(B) = s$, $q_{12}(B) > s$ $(n > 2, s > 0, p = 2)$.*

PROOF: If we transform $W_{21}AW_{21}^{-1}A$ into pseudo-normal form according to the method of lemma 4.4, we shall have the required matrix.

In the remaining case $q_{12}(A) = r_{12}(A) = s(A) = s$; $r_{32}(A) > s$, we may assume without loss of generality that $a_{13} = 0$. (Otherwise transform by W_{23}^m.)

LEMMA 7.3. *If $A = (a_{ij})$ is in pseudo-normal form and $q_{12}(A) = r_{12}(A) = s(A) = s > 0$; $r_{32}(A) > s$; $a_{13} = 0$; $r_{23}(A) = s$, or $r_{43}(A) = s$, $n \geqq 4$; then $\mathfrak{S}_A$ contains a matrix B in p.n. form with $s(B) = s$; $q_{12}(B) > s$ $(n > 2, s > 0, p = 2)$.*

PROOF: Without loss of generality assume that $r_{23}(A) = s$. (If necessary transform by W_{24}.) Consider $C = A \cdot W_{31}^{-1}AW_{31}$. We have $q_{12}(C) > s$, $r_{21}(C) = s$, $r_{51}(C) = r_{61}(C) = \cdots = r_{n1}(C) = r$, $r_{23}(C) > s$, $r_{24}(C) > s$. According to the method of lemma 4.4, C may be transformed into p.n. form by first transforming with $W_{32}^m W_{42}^{m'}$, and by subsequent transformations which do not affect $q_{12}(C)$. Hence when C is transformed by this process into p.n. form, it becomes the required B.

LEMMA 7.4. *If $A = (a_{ij})$ is in p.n. form; $q_{12}(A) = s(A) = s > 0$; $r_{12}(A) = s$; $r_{32}(A) > s$; $a_{13} = 0$; $r_{23}(A) > s$, $n < 4$; or $r_{43}(A) > s$, $n \geqq 4$; then $\mathfrak{S}_A$ contains a matrix B in p.n. form with $s(B) = s$, $q_{12}(B) > s$ $(n > 2, s > 0, p = 2)$.*

PROOF: Consider $C = W_{31}AW_{31}^{-1}$. The first step in transforming C into p.n. form is transformation by W_{32}^t $(t(2^s + a_{23}) = a_{33} - a_{11}\ (2^r))$. Since either $a_{33} - a_{11} = 0\ (2^{s+1})$ or else $a_{33} = a_{22}\ (2^{s+1})$, it is easy to see that $r_{32}(W_{32}^t C W_{32}^{-t}) = s$. Thus if C be transformed into p.n. form according to the method of lemma 4.4, the resulting matrix B will satisfy $q_{12}(B) = r_{32}(B) = s$. See lemma 7.1.

LEMMA 7.5. *If $n > 2$, $s > 0$, $p = 2$, and A is in pseudo-normal form and $q_{12}(A) > s(A) = s > 0$, then $\mathfrak{S}_A$ contains a matrix B with $q_{12}(B) > s$, $r_{21}(B) = s$; $r_{ij}(B) = r$ $(j = 1, 2; i = 3, 4, \cdots, n)$.*

PROOF: If we use the notation of lemma 6.2 for α, m, then in the matrix $D = (d_{ij}) = V_1^{(\alpha)} A V_1^{(1/\alpha)} W_{12}^m A W_{12}^{-m}$, if we set $d_{12} = l2^s$, then l is odd; $d_{11} = d_{22}\ (2^{s+1})$; $a_{22} + a_{33} = k$ is even. Permute[7] the subscripts 1 and 2; if $2a_{32} = 0\ (2^r)$ there is nothing more to prove. If not, we form

$$F = (f_{ij}) = W_{32}^m W_{42}^{m'} E W_{42}^{-m'} W_{32}^{-m},$$

where $ml2^s - ka_{32} = m'l2^s - a_{43}a_{32} = 0\ (2^r)$; we can verify that $\min_{\theta>2} r_{\theta 2}(F) > r_{32}(A) = (a_{32}, 2^r)$. If therefore F is transformed into pseudo-normal form according to the method of lemma 4.4, we shall reach a matrix A' like A, only with $r_{32}(A') > r_{32}(A)$. Continuing in this way, we reach B eventually.

LEMMA 7.6. *Let $B = (b_{ij})$ be as defined in the previous lemma. Then $\mathfrak{S}_B$ contains a matrix C in normal form, with $r_{12}(C) = s(B) = s$, $q_{12}(C) > s$, and with the lower right $(n-2)^2$ box of C the $(n-2)^2$ unit matrix.*

PROOF: $V_1^{(\alpha)} B V_1^{(1/\alpha)} W_{12}^m B^{-1} W_{12}^{-m}$, where α and m satisfy $\alpha = 1 + 2^s$, $m = (1/\alpha)b_{11}\ (2^r)$, is the matrix C with the required properties.

LEMMA 7.7. *Let $C = (c_{ij})$ be as defined in the previous lemma. Then $\mathfrak{S}_C$ contains the matrix $D' = (D, \mathbf{0}, \mathbf{0}, \mathbf{1})$,[6] where $D = (1 + 2^s, 2^s, 0, 1/(1 + 2^s))$.*

PROOF: Assume without loss of generality that $c_{13} = c_{14} = \cdots = 0$. (Otherwise transform by $W_{23}^{m_3} W_{24}^{m_4} \cdots W_{2n}^{m_n}$.) The matrix $E = (e_{ij}) = W_{21} C W_{21}^{-1} C^{-1}$ is an open product with $r_{21}(E) = s$; $q_{12}(E) > s$; $E = (F, \mathbf{0}, \mathbf{0}, \mathbf{1})$, so by lemmas 6.2, 6.6, $\mathfrak{S}_E$ contains D' of this lemma.

LEMMA 7.8. *The least normal subgroup containing D' of the previous lemma is $\mathfrak{N}_s$.*

PROOF: $W_{31}^{-1} D' W_{31} D'^{-1} = W_{31}^w\ (w = 2^s)$. See theorem 1.5.

THEOREM 7.9. *When $n > 2$, $s(A) > 0$, $p = 2$, then $\mathfrak{S}_A$ contains $\mathfrak{N}_{s(A)}$.*

8. Fundamental theorem for $n = 2$, $s = 0$, $p = 2$.

LEMMA 8.1. *If $p = 2$, $n = 2$, $s(A) = 0$, then $\mathfrak{S}_A$ contains $\mathfrak{J}_1$ (and hence $\mathfrak{N}_2$).*

PROOF: Assume without loss of generality that A is in canonical form (lemma 3.1). Set $m = a_{22} - 2$. Then $a_{22} + m$ is even. Consider $B = W_{12}^m A W_{12}^{-m} \cdot A = (a_{12}, 2, a_{12}(a_{22} + m), a_{12} + 2(a_{22} + m))$.[6] Since $s(B) = 1$, $q_{12}(B) > 1$, $\mathfrak{S}_B$ contains $\mathfrak{J}_1$ by theorem 3.7.

LEMMA 8.2. *If $p = 2$, $n = 2$, $s(A) = 0$, then $\mathfrak{S}_A$ contains $(0, 1, -1, -1)$.*

PROOF: Assume without loss of generality that A is in canonical form (lemma 3.1). Set $m = a_{12} + a_{22} - 1$, $\alpha = -1/a_{12}$. Consider the process of forming B from A, where

$$B = V_1^{(\alpha)} W_{12}^m A W_{12}^{-m} W_{12}^{-1} A W_{12} V_1^{(1/\alpha)}$$
$$= V_1^{(\alpha)}(0, -a_{12}, a_{12}, a_{12}(2 - 2a_{22} - a_{12})) V_1^{(1/\alpha)} = (0, 1, -a_{12}^2, b_{22}).$$

If this process be repeated on B, we reach $C = (0, 1, -a_{12}^4, c_{22}) \cdots$, and eventually $D = (0, 1, -1, d_{22})$ where $d_{22} = -1\ (4)$. Now $-1/d_{22} = 1\ (4)$, so that $E = (-d_{22}, 0, 0, -1/d_{22})$ is in $\mathfrak{N}_2$ and hence in $\mathfrak{S}_A$. If $\alpha = -d_{22}$, $\mathfrak{S}_A$ contains also $V_1^{(\alpha)} D E V_1^{(1/\alpha)} = (0, 1, -1, -1)$, Q.E.D.

THEOREM 8.3. *If $s(A) = 0$, the matrices of $\mathfrak{S}_A$ of determinant 1 (2^r) fill out $\mathfrak{N}_0$, $\mathfrak{F}$ or $\mathfrak{K}$.* See lemmas 5.4, 5.5.

9. **Fundamental theorem for $n > 2$, $s = 0$ $(p \geqq 2)$.**

LEMMA 9.1. *If $s(B) = 0$, B has a conjugate $A = (a_{ij})$ in pseudo-normal form such that $a_{1j} = 0$ (p^r) $(j \neq 2)$.*

PROOF: By lemmas 3.1, 4.4, B has a conjugate $C = (c_{ij})$ in p.n. form with $c_{11} = 0$ (p^r). We may assume without loss of generality that $c_{12} \not\equiv 0$ (p). (Otherwise $c_{32} \not\equiv 0$ (p) and we can transform by W_{13}.) Then A is the conjugate obtained from C by transforming successively by $W_{2j}^{m_j}$ $(j = 3, 4, \cdots, n;$ $m_j c_{12} \equiv c_{1j}$ $(p^r))$.

LEMMA 9.2. *Let $s(A) = 0$, and let $A = (a_{ij})$ be in pseudo-normal form with $a_{1j} = 0$ (p^r) $(j \neq 2)$. Then $\mathfrak{S}_A$ contains a matrix $D = (d_{ij})$ with $d_{21} \not\equiv 0$ (p) and such that the last $n - 2$ rows of D are identical with those of the unit matrix.*

PROOF: Set $A^{-1} = A^* = (a_{ij}^*)$. Then since $A^*A = \mathbf{1}$, we see (on computing the element in the 1, 1th place in A^*A) that $a_{21}^* \not\equiv 0$ (p). Then $D = A^{-1}W_{12}^{-1}AW_{12}$.

LEMMA 9.3. *Let D be as defined in lemma 9.2. Then $\mathfrak{S}_D$ contains an open product $F = (E, \mathbf{0}, \mathbf{0}, \mathbf{1})$,[6] with $s(E) = 0$.*

PROOF: Without loss of generality, assume that d_{12} is prime to p, and that $d_{1j} = 0$ (p^r) $(j \neq 2)$. (Lemmas 3.1, 9.1.) Then $W_{21}DW_{21}^{-1}D^{-1} = F$. In fact, $r_{12}(F) = 0$.

THEOREM 9.4. *When $p = 2$, $n > 2$, $s(B) = 0$, then $\mathfrak{S}_B \supseteq \mathfrak{N}_0$.*

PROOF: Because of the lemmas of this section and lemma 8.2, we may assume without loss of generality that B is $(B', \mathbf{0}, \mathbf{0}, \mathbf{1})$, where $B' = (0, 1, -1, -1)$. Let $B^2W_{31}^{-1}BW_{31} = H$. By lemma 8.1, $\mathfrak{S}_B \supset K = W_{31}^2$ diag $(-1, 1, -1, \mathbf{1})$. Call L the matrix obtained by permuting[7] 1 and 2 in $HKV_2^{(-1)}HKV_2^{(-1)}$; then $HL = W_{32}^{-1}$, which shows that $\mathfrak{S}_B = \mathfrak{N}_0$.

THEOREM 9.5. *When $p \geqq 2$, $n > 2$, $s(A) = 0$, then $\mathfrak{S}_A \supseteq \mathfrak{N}_0$.*

10. **The normal subgroups when $p > 2$. Summary.** Now let any normal subgroup $\mathfrak{N}$ of $\mathfrak{G}$ be given. We propose to give an algorithm for telling what matrices $\mathfrak{N}$ contains when $p > 2$.

We define $s(\mathfrak{N}) = \min s(A)$, $A \in \mathfrak{N}$. The multiplier (p^s) of a matrix $A \in \mathfrak{N}$ is the residue class mod p^s to which all the elements of the main diagonal of A are congruent (mod p^s) $(s = s(\mathfrak{N}))$. We define the first associated subgroup $\mathfrak{N}'$ of $\mathfrak{N}$ as the subgroup of $\mathfrak{N}$ containing its matrices of determinant $1(p^r)$. The second associated subgroup $\mathfrak{N}''$ is defined as the subgroup of $\mathfrak{N}'$ all of whose matrices have multipliers $1(p^s)$. $(\mathfrak{N}')' = \mathfrak{N}'$; $(\mathfrak{N}')'' = (\mathfrak{N}'')' = (\mathfrak{N}'')'' = \mathfrak{N}_{s(\mathfrak{N})}$.[5a]

The multiplier (p^s) of the product of two matrices is congruent to the product of their multipliers (p^s). Therefore the multipliers (p^s) $(s = s(\mathfrak{N}))$ of any normal subgroup $\mathfrak{N}$ of $\mathfrak{G}$ form a group 1, 1 isomorphic with a subgroup of the multiplicative group (p^s) of reduced residue classes, and the same is true of $\mathfrak{N}'$, since $s(\mathfrak{N}') = s(\mathfrak{N})$ because $\mathfrak{N}' \supseteq \mathfrak{N}_s$.

THEOREM 10.1. *The quotient group $\mathfrak{N}'/\mathfrak{N}''$ is cyclic* ($p > 2$).

In fact, as generator of this cyclic group we may take a coset containing a matrix whose multiplier (p^s) is a generator of the corresponding homorphic cyclic group of residue classes (p^s).

Let numbers representative of the homorphic cyclic group of residue classes (p^s) be

$$\gamma, \gamma^2, \cdots, \gamma^l = 1 \ (p^s). \tag{10.2}$$

THEOREM 10.3. *A matrix A is in $\mathfrak{N}'$ if and only if $s(A) \geqq s(\mathfrak{N}')$, $\det A = 1 \ (p^r)$, and the multiplier (p^s) of A is one of the numbers* (10.2) ($p > 2$).

THEOREM 10.4. *The quotient group $\mathfrak{N}/\mathfrak{N}'$ is cyclic* ($p > 2$).

It is in fact isomorphic with that group of reduced residue classes mod p^r whose elements are the various determinants of matrices of $\mathfrak{N}$.

One way of determining whether a given matrix A is in $\mathfrak{N}$ is the following. Take any matrix B of determinant Δ from a generating coset of $\mathfrak{N}/\mathfrak{N}'$, and form $B, B^2, \cdots, B^t$, where t is the exponent to which Δ belongs (p^r). Then if A has determinant Δ^k, form $C = AB^{t-k}$ and decide whether C lies in $\mathfrak{N}'$. Or alternatively, one may use the following criterion, which amounts to the same thing, and is worth stating as

THEOREM 10.5. *If $p > 2$, the matrix A is in $\mathfrak{N}$ if and only if: $s(A) \geqq s(\mathfrak{N})$; $\det A$ is one of the numbers $\Delta, \Delta^2, \cdots, \Delta^t = 1 \ (p^r)$, say Δ^j; the multiplier ζ (p^s) of A satisfies $\zeta = \gamma^k \eta^j$ for some k, where η is the multiplier (p^s) of some fixed matrix of determinant Δ of $\mathfrak{N}$.*

We have found all the normal subgroups of $\mathfrak{G}$.

Further remarks about the multipliers (p^s) of $\mathfrak{N}'$ are in order. We have for each γ^k in the set (10.2): $(\gamma^k)^n = 1 \ (p^s)$. Hence if $(n, \phi(p^s)) = 1$, then $\gamma^k = 1 \ (p^s)$, and $\mathfrak{N}' = \mathfrak{N}''$.

THEOREM 10.6. *$\mathfrak{N}' = \mathfrak{N}''$ whenever $(n, p(p-1)) = 1$, as well as when $s(\mathfrak{N}') = 0$ ($p > 2$).* For when $s(\mathfrak{N}') = 0$, $\mathfrak{N}'' \supseteq \mathfrak{N}'$.

If more generally $(n, \phi(p^s)) = \delta$, then $(\gamma^k)^\delta = 1 \ (p^s)$. This limits the possible order of $\mathfrak{N}'/\mathfrak{N}''$ to divisors of δ, and p never divides this order more than (δ, p^{s-1}) times ($s > 0$), or in general we may state

THEOREM 10.7. *If $(n, p^{r-1}) = p^u$, then p never divides the order of $\mathfrak{N}'/\mathfrak{N}''$ more than* min $(u, s-1)$ *times ($s = s(\mathfrak{N}) > 0$). If $(n, \phi(p^r)) = d$, then the order of $\mathfrak{N}'/\mathfrak{N}''$ is a divisor of $p^{-u}d$* min (p^u, p^{s-1}) ($p > 2$).

We have here discussed the first associated groups $\mathfrak{N}'$, with $s(\mathfrak{N}') = s$ which include a given $\mathfrak{N}_s$. To finish this discussion we state the obvious

THEOREM 10.8. *There is either one or no first associated group $\mathfrak{M}'$ with $s(\mathfrak{M}') = s > 0$ containing a given $\mathfrak{N}_s$ as a subgroup with preassigned index q, according as q is or is not a divisor of $p^{-u}d$* min (p^u, p^{s-1}), *where $(n, \phi(p^r)) = d$, $(d, p^r) = p^u$. —If $s = 0$, then $\mathfrak{M}'$ is always the same as $\mathfrak{N}_s$ ($p > 2$).*

We now discuss similarly the multipliers (p^s) of $\mathfrak{N}$, any normal subgroup of $\mathfrak{G}$. Let the various determinants of the matrices of $\mathfrak{N}$ be $\Delta, \Delta^2, \cdots, \Delta^t = 1 \ (p^r)$; we know that a Δ exists so that they can each be written in one of these

forms. Let Δ belong mod p^s to the exponent t'. The theory of primitive roots mod p^r shows that $t' = \phi(p^s)/[\phi(p^r)/t, \phi(p^s)]$.[3] Let B be a matrix of $\mathfrak{R}$ of determinant Δ, and let its multiplier be η (p^s). Then $\eta^n = \Delta$ (p^s). Hence if $(n, \phi(p^s)) = 1$, η belongs mod p^s to the exponent t', and is furthermore determined mod p^s by Δ.

THEOREM 10.9. *If $(n, \phi(p^s)) = 1$, then the order of $\mathfrak{R}/\mathfrak{R}'$ is a divisor of $\phi(p^r)$, and if $\mathfrak{R}'$ be preassigned, then for every divisor t of $\phi(p^r)$ there is just one normal subgroup $\mathfrak{M}$ with $s(\mathfrak{M}) = s(\mathfrak{R}') = s$ having $\mathfrak{R}'$ for a first associated and containing $\mathfrak{R}'$ with index t $(p > 2)$.*

More generally, let $(n, \phi(p^s)) = \delta$. Then η^δ is determined mod p^s from Δ, but in various cases there might be several solutions η_1 to (10.10) $\eta_1^\delta = \eta^\delta$ (p^s). In fact, if ϵ is any solution of (10.11) $\epsilon^\delta = 1$ (p^s), then $\eta_1 = \eta\epsilon$ satisfies (10.10), and conversely if η_1 satisfies (10.10) then $\eta_1\eta^{-1}$ satisfies (10.11). The number of solutions of (10.11) is easily shown to be δ by the theory of primitive roots mod p^s.

THEOREM 10.12. *Let a normal subgroup $\mathfrak{R}'$ imbedded in the commutator group be given; set $s(\mathfrak{R}') = s$, and let an integral divisor t of $\phi(p^r)$ be given. Then necessary and sufficient conditions that there exist a normal subgroup $\mathfrak{M}$ (with $s(\mathfrak{M}) = s$) having $\mathfrak{R}'$ for a first associated, with $\mathfrak{M}/\mathfrak{R}'$ of order t, are that*

a) $\delta \mid \phi(p^s)/t'$, *where* $\delta = (n, \phi(p^s))$; $t' = \phi(p^s)/[\phi(p^r)/t, \phi(p^s)]$;

b) $[t', (\delta/l)/(\delta/l, t/t')] = 1$,[3] *where l is the order of $\mathfrak{R}'/\mathfrak{R}_s$. Furthermore, the number of such groups $\mathfrak{M}$, if there are any, is $(\delta/l, t/t')$ $(p > 2)$.*

PROOF: NECESSITY. Let $\mathfrak{M}$ be given having $\mathfrak{R}'$ for a first associated, with $s = s(\mathfrak{M}) = s(\mathfrak{R}')$; $\mathfrak{M}/\mathfrak{R}'$ of order t. The determinants of the matrices of $\mathfrak{M}$ are say the numbers $\Delta, \Delta^2, \cdots, \Delta^t = 1$ (p^r). Let the multiplier (p^s) of some matrix B of $\mathfrak{M}$ of determinant Δ be η. Then $\eta^n = \Delta$ (p^s). Since $(n, \phi(p^s)) = \delta$, the exponent to which η^n belongs mod p^s is a divisor of $\phi(p^s)/\delta$; the exponent to which Δ belongs mod p^s is t'. Hence $t' \mid \phi(p^s)/\delta$, $\delta \mid \phi(p^s)/t'$, which proves a).

The solutions of $\eta^n = \Delta$ (p^s) are the same as the solutions of

$$\eta^\delta = \Delta^x \quad (p^s), \tag{10.13}$$

where we have set $n = \delta k$, and we have solved $\delta k x = \delta$ $(\phi(p^s))$, a solution of which may be found by solving $kx = 1$ $(\phi(p^s)/\delta)$, which in turn is soluble because $(k, \phi(p^s)/\delta) = (n/\delta, \phi(p^s)/\delta) = 1$ from the definition of δ. It is easy to see that Δ and Δ^x belong mod p^s to the same exponent. In fact, Δ^x belongs mod p^s to the exponent $t'/(x, t')$. But since $t' \mid \phi(p^s)/\delta$ and $(x, \phi(p^s)/\delta) = 1$, then $(x, t') = 1$. Thus η^δ belongs mod p^s to the exponent t'. We now set up all solutions of (10.13) using the fact that if f belongs mod p^s to the exponent t', then there is a primitive root ϵ mod p^s such that $\epsilon^{u/t'} = f$, $u = \phi(p^s)$ $(p > 2)$. (This is proved in the abstract theory of cyclic groups.)

Let us write Δ^x in the form $\epsilon^{u/t'}$ (p^s), where ϵ is a primitive root mod p^s. Set $\phi(p^s) = h\delta t'$. Then $\Delta^x = \epsilon^{h\delta}$ (p^s), while all the solutions ϵ_1 of $\epsilon_1^\delta = 1$ (p^s) are $\epsilon_1 = \epsilon^{iht'}$, $i = 1, 2, \cdots, \delta$. Hence all the solutions ξ of $\xi^\delta = \Delta^x$ (p^s) are the numbers $\xi = \epsilon^{h+iht'}$, $i = 1, 2, \cdots, \delta$. Hence η is one of these numbers, say

$\eta = \epsilon^{h+i_1 ht'}$. Now since B^t is in $\mathfrak{N}'$, $(\eta^t)^l = 1\ (p^s)$. Hence if we set $(\delta/l, t/t') = g$, we have $1 + i_1 t' = 0\ (\delta/lg)$, which makes it obvious that $(t', \delta/lg) = 1$.

SUFFICIENCY. Firstly, if the condition $\delta \mid \phi(p^s)/t'$ is satisfied, then $t' \mid \phi(p^s)/\delta$, so that we can solve $\eta^n = \Delta\ (p^s)$. Then we can determine λ in the matrix $B = \text{diag}\ (\eta + \lambda p^s, \eta, \eta, \cdots, \eta)$ so that this matrix has determinant Δ. Then the matrix B together with $\mathfrak{N}'$ will generate an $\mathfrak{M}$ with the required properties. In particular, we can show that an η can be found to satisfy $(\eta^t)^l = 1\ (p^s)$ by retracing steps in the last part of the proof on necessity.

We have also to count the number of distinct normal subgroups with the specified properties. To this end let us find the number of solutions η of $\eta^\delta = \Delta^x\ (p^s)$ which also satisfy $\eta^{lt} = 1$. Basing our investigation on the same primitive root ϵ that was used in the necessity proof, we observe that the desired numbers η are the numbers $\epsilon^{h(1+it')}\ (p^s)$, where i is to be specified mod δ, and i satisfies $1 + it' = 0\ (\delta/lg)$, which has lg solutions i. But in each group $\mathfrak{M}$ there are l distinct multipliers (p^s) among the matrices of determinant Δ, and a consideration of the group of reduced residue classes mod p^s shows that the lg solutions of the above congruence are thus separated into g sets of l each. Hence there are $g = (\delta/l, t/t')$ groups $\mathfrak{N}$ with the specified properties. They are all isomorphic.

Summary. A brief summary of the results of 10 is advisable at this point. Subsequently these results are applied to the case $r = 1$, which holds especial interest. The normal subgroups of $\mathfrak{G}$ have been classified according to four natural parameters. First we have $s = s(\mathfrak{N})$, which may be any number from 0 to r inclusive. Next we have l, the order of $\mathfrak{N}'/\mathfrak{N}_s$, which may be any divisor of $\delta = (n, \phi(p^s))$. Next we have t, the order of $\mathfrak{N}/\mathfrak{N}'$. The three restrictions for t are (a) $t \mid \phi(p^r)$; (b) $\delta \mid \phi(p^s)/t'$; (c) $(t', \delta/lg) = 1$, where

$$t' = \phi(p^s)/[\phi(p^r)/t, \phi(p^s)], \qquad g = (\delta/l, t/t').$$

Lastly, we have the fourth parameter, by means of which we number the g groups whose first three parameters are identical. This numbering can be done in ultra-specific form if we pick for Δ a special number belonging mod p^r to the exponent t, say the least positive one. Then ϵ is determined as the least positive integer satisfying $\epsilon^{u/t'} = \Delta^x\ (p^s)$, $u = \phi(p^s)$, and the numbering can be carried through by another specific convention on numbering the powers of ϵ which satisfy $\eta^{lt} = \epsilon^{lt(h+iht')} = 1(p^s)$.

In case $n = 1$, $\mathfrak{G}$ is simply isomorphic with the group of automorphisms of a cyclic group of order p^r. The group $\mathfrak{G}$ is itself cyclic, and one parameter is necessary to specify its normal subgroups (namely the order).—In case $(n, \phi(p^r)) = 1$, the second and fourth parameters are always 1, and the third parameter t must satisfy only $t \mid \phi(p^r)$.—If especially $r = 1$, then $\mathfrak{G}$ is the group of automorphisms of the elementary group (direct product of n cyclic groups of order p). In this case, $\mathfrak{N}_0 = \mathfrak{C}_0$, and $\mathfrak{N}_r = \mathfrak{N}_1 = \mathbf{1}$, the identity.

THEOREM 10.15. *When $\mathfrak{G}$ is the group of automorphisms of the elementary group (i.e. when $r = 1$), then the normal subgroups of $\mathfrak{G}$ which are imbedded in the commutator group are all characteristic ($p > 2$).*

PROOF: The groups in question are the identity, the whole commutator group, and the normal subgroups $\mathfrak{N}'$ with $s(\mathfrak{N}') = 1$; the order of such a group is the exponent mod p to which its generating multiplier γ belongs, and is a divisor of $(n, p-1)$. (Theorem 6.7.) Since the groups are the only normal subgroups which are also subgroups of the characteristic subgroup $\mathfrak{C}_0$, and since they all have different orders, they must all be characteristic. Incidentally, we have counted them.

THEOREM 10.16. *When $\mathfrak{G}$ is the group of automorphisms of the elementary group (i.e. when $r = 1$), then if $(n, p-1) = 1$, the normal subgroups of $\mathfrak{G}$ are all characteristic ($p > 2$).*

PROOF: They can again be classified by order (theorem 10.12).

11. **The normal subgroups when $n = 2$, $p = 2$.**[8]

First, let $s > 0$. From 5, 6, we have $\mathfrak{N}'' = \mathfrak{N}_s$, $\mathfrak{J}_s$, or $\mathfrak{H}_{s+1}$ or $\mathfrak{L}_2$ or $\mathfrak{P}_2$, $s = s(\mathfrak{N})$.

Note that $\mathfrak{H}''_{s+1} = \mathfrak{N}_{s+1}$; $\mathfrak{L}''_2 = \mathfrak{P}''_2 = \mathfrak{J}_2$. Hence in theorem 11.1, $\mathfrak{N}'' = \mathfrak{N}_s$ or $\mathfrak{J}_s$.

THEOREM 11.1. *Let $\mathfrak{N}''$ be given, $s = s(\mathfrak{N}'') > 0$. A necessary and sufficient condition that the first associated group $\mathfrak{M}'$ have $\mathfrak{N}''$ for a second associated, $s(\mathfrak{M}') = s$, is: if $s = 1$: $\mathfrak{M}' = \mathfrak{N}''$; if $s = 2$: $\mathfrak{M}' = \mathfrak{N}''$, or $\mathfrak{M}' = \{\mathfrak{N}'', -\mathbf{1}\}$ or $\mathfrak{P}_2$; if $s \geqq 3$: $\mathfrak{M}' = \mathfrak{N}''$, or $\mathfrak{M}' = \{\mathfrak{N}'', -\mathbf{1}\}$, or possibly (if $\mathfrak{N}'' = \mathfrak{N}_s$), $\mathfrak{M}' = \{\mathfrak{N}'', (-1 + 2^{s-1}, 0, 0, 1/(-1 + 2^{s-1}))\}$ or $\mathfrak{H}_s$ or $\{\mathfrak{H}_s, -\mathbf{1}\}$.*[6]

REMARK: All these groups contain $\mathfrak{N}''$ with index 1, 2, 4 and they are all distinct.

PROOF: A) By definition of $\mathfrak{M}'$, $(\mathfrak{M}')'' = \mathfrak{N}'$, $(\mathfrak{M}')' = \mathfrak{M}'$; $\mathfrak{M}'$ is normal.

B) Let the matrix $A = (a_{ij})$ be in a group $\mathfrak{M}'$ with $s(\mathfrak{M}') = s(\mathfrak{N}'')$, $(\mathfrak{M}')'' = \mathfrak{N}''$. Then $\det A = a_{11}^2 = 1\ (2^s)$, hence $a_{11} = \pm 1\ (2^s)$ or possibly (if $s \geqq 3$) $a_{11} = \pm 1 + 2^{s-1}\ (2^s)$. But the condition $(\mathfrak{M}')'' = \mathfrak{N}''$ eliminates $a_{11} = 1 + 2^{s-1}\ (2^s)$ as well as $-1 + 2^{s-1}\ (2^s)$ when $\mathfrak{N}'' = \mathfrak{J}_s$, for A^2 is in $\mathfrak{N}_s$ (but not in $\mathfrak{J}_s$) in this case. Thus we reach the groups $\mathfrak{M}'$ of the theorem.

LEMMA 11.2. *If $s = 2$ or 3, then $r_{21}(A) = r_{12}(A) = q_{12}(A) = s - 1$ is impossible if $\det A = 1\ (2^r)$ ($r \geqq s$, $n = 2$).*

PROOF: Suppose the contrary. Then $a_{22}(a_{11} - a_{22})/2^{s-1}$ must be odd, while $(a_{11}a_{22} - 1)/2^{s-1}$ must be even since $\det A - 1 = 0\ (2^r)$. Hence their difference $(1 - a_{22}^2)/2^{s-1}$ is odd, which is impossible since $a_{22}^2 = 1\ (8)$.

THEOREM 11.3. *Let $\mathfrak{N}''$ be given, $s = s(\mathfrak{N}'') > 0$. A necessary and sufficient condition that the first associated group $\mathfrak{M}'$ have $\mathfrak{N}''$ for a second associated is:*

if $\mathfrak{N}'' = \mathfrak{J}_s$ or $\mathfrak{N}_s$, or if $s < 4$: $\mathfrak{M}'$ be as in theorem 11.1;

if $s \geqq 4$ and $\mathfrak{N}'' = \mathfrak{H}_s$: $\mathfrak{M}' = \mathfrak{N}''$ or $\{\mathfrak{N}'', -\mathbf{1}\}$ or $\{\mathfrak{N}'', M\}$ or $\{\mathfrak{N}'', -M\}$ or $\{\mathfrak{N}'', M, -\mathbf{1}\}$, where $M = (1 + 2^{s-2}, 2^{s-1}, 2^{s-1}, (1 + 2^{2(s-1)})/(1 + 2^{s-2}))$.

REMARK: If $(\mathfrak{M}')' = \mathfrak{M}'$, $(\mathfrak{M}')'' = \mathfrak{N}''$, then $o\{\mathfrak{M}'/\mathfrak{N}''\} = 1$ or 2 except in the last case, when $\mathfrak{M}'/\mathfrak{N}''$ is of type (1, 1).

---

[8] Any group mentioned is a normal subgroup of $\mathfrak{G}_{2,2,r}$, unless specifically described otherwise.

PROOF: Theorem 11.1 does not apply only when $s(\mathfrak{M}') < s(\mathfrak{N}'')$. In this case (§6) $\mathfrak{N}'' = \mathfrak{H}_s$ and $\mathfrak{M}'$ contains a matrix $A = (a_{ij})$ with $r_{21}(A) = r_{12}(A) = q_{12}(A) = s - 1$. Hence by lemma 11.2, $s \geqq 4$; if B is another matrix with $B \in \mathfrak{M}'$ and $s(B) = s - 1$, then $r_{21}(B) = r_{12}(B) = q_{12}(B) = s - 1$. But the matrices A satisfying such a restriction fall into two cosets mod $\mathfrak{H}_s$, represented by $\pm M$.

For we have $a_{11} = \pm 1$ or $\pm 1 + 2^{s-2}$ (2^{s-1}), hence $a_{11} = \pm 1$ or $\pm 1 + 2^{s-1}$ or $\pm 1 + 2^{s-2}$ or $\pm 1 - 2^{s-1}$ (2^s), and moreover $a_{12} = a_{21} = 2^{s-1}$ (2^s). Mod $\mathfrak{H}_s$ the values $a_{11} = \pm 1$ and $\pm 1 + 2^{s-1}$ are the same, as are $a_{11} = \pm 1 + 2^{s-2}$ and $\pm 1 - 2^{s-2}$. Of the four matrices mod $\mathfrak{H}_s$ remaining, $(\pm 1, 2^{s-1}, 2^{s-1}, \pm (1 + 2^{2(s-1)}))$ do not have $q_{12} = s - 1$, while $\pm M$ satisfy all requirements, Q.E.D.

We have given a complete description of the groups $\mathfrak{N}'$, $(s(\mathfrak{N}') > 0)$; since we know representatives of each coset of $\mathfrak{N}'/\mathfrak{N}_s$ or $\mathfrak{N}'/\mathfrak{N}_{s+1}$, $s = s(\mathfrak{N}')$, we could give an algorithm for telling (by multiplying A by each of three matrices) whether a given matrix A is in $\mathfrak{N}'$.

THEOREM 11.4. *If $s = s(\mathfrak{N}) \geqq 3$, $\mathfrak{N}/\mathfrak{N}'$ is cyclic and its order divides 2^{r-3}. If $s = 2$, $\mathfrak{N}/\mathfrak{N}'$ is cyclic and its order divides 2^{r-2}. If $s = 1$, $\mathfrak{N}/\mathfrak{N}'$ is the direct product of a cyclic group whose order divides 2^{r-2} by a subgroup of the two group.*

PROOF: The determinants of matrices in two distinct cosets of $\mathfrak{N}/\mathfrak{N}'$ are distinct mod 2^r. The various determinants must moreover be quadratic residues mod 2^s. See lemma 12.2. When $s = 2$, we observe that -3 is congruent to a residue mod 2^s, while 3 is not.

If $s > 2$, the determinants of $\mathfrak{N}$ are a subset of 3^{2i} $(i = 1, \cdots, 2^{r-3})$. If $s = 2$ the determinants of $\mathfrak{N}$ are a subset of $(-3)^i$ $(i = 1, \cdots, 2^{r-2})$.

THEOREM 11.5. *Let the first associated group $\mathfrak{N}'$ be given, $s = s(\mathfrak{N}') > 0$. Necessary and sufficient conditions that there exist a normal subgroup $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}'$, $s(\mathfrak{M}) < s$, $\mathfrak{M}/\mathfrak{N}'$ of preassigned order u, are: if $s = 2$: $u = 2^\gamma$ $(0 < \gamma < r)$; if $s = 3$: $u = 2^{r-2}$; if $s \geqq 4$: $u = 2^\gamma$ $(0 < \gamma < r - 2)$. When u satisfies these conditions, then if $s = r = 2$ there is one subgroup $\mathfrak{M}$; if $s = 2$, $r \geqq 3$ there are two such groups $\mathfrak{M}$ when $u = 2$ one otherwise $(\mathfrak{N}' = \mathfrak{H}_2)$; if $s \geqq 4$ there are one or two groups $\mathfrak{M}$, according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$.*

PROOF: NECESSITY. By theorems 6.9, 8.3, $s(\mathfrak{M}) = s - 1 > 0$, and $(\mathfrak{N}')'' = \mathfrak{H}_s$ or $\mathfrak{L}_2$ or $\mathfrak{P}_2$.

Let $s = 2$ or 3. As in lemma 11.2, if A is in $\mathfrak{M}$ but not in $\mathfrak{N}'$, $(\det A - 1)/2^{s-1}$ must be odd. If $s = 2$, then $\det A = -1$ (8) or possibly (if $r > 2$) $\det A = 3$ (8). If $\det A = -1$ (8), then A mod $\mathfrak{H}_2 = N = (1, 2, 2, 4 - 3^j)$, $j = 2^{r-2-\gamma}$ $(0 \leqq \gamma < r - 2)$; mod $\mathfrak{L}_2$ or $\mathfrak{P}_2$ we must consider also $N' = (-1, 2, -2, 4 + 3^j)$ or $N'' = -N$ respectively; when $r = 2$, $N = (1, 2, 2, 3)$. The matrix N is of order 2 mod $\mathfrak{N}'$ if $\gamma = 0$; it is of order 2^γ mod $\mathfrak{N}'$ if $0 < \gamma < r - 2$. If $\det A = 3$ (8), then $M = (1, 2, 2, -1)$ generates A mod $\mathfrak{H}_2$ and is generated by A mod $\mathfrak{H}_2$; mod $\mathfrak{L}_2$ or $\mathfrak{P}_2$ we must consider also $M' = (-1, 2, -2, 1)$ or $M'' = -M$ respectively; M is of order 2^{r-2} mod $\mathfrak{N}'$. We have $\mathfrak{M} = \{\mathfrak{N}', N\}$ or $\{\mathfrak{N}', M\}$, or $\{\mathfrak{N}', M, N\}$. If $s = 3$ then $\det A = -3$ (8), and $M = (1, 4, 4, 13)$ generates A mod $\mathfrak{N}'$ and is generated by A mod $\mathfrak{N}'$; M is of order 2^{r-2} mod $\mathfrak{N}'$. We have $\mathfrak{M} = \{\mathfrak{N}', M\}$.

When $s \geqq 4$, by theorem 11.4 $\mathfrak{M} = \{\mathfrak{N}', A\}$, $s(A) = s - 1$. Set $A = (a_{ij})$, then $a_{11}^2 = \det A = 3^j\ (2^{s-1})$. Assume without loss of generality that $j = 2^\gamma$, $0 < \gamma < r - 2$. Solving this quadratic congruence we find $a_{11} = \pm 3^{j/2}$ or $\pm 3^{j/2}(1 + 2^{s-2})\ (2^{s-1})$. Let us abbreviate by t that one of these numbers to which a_{11} is congruent (2^{s-1}). Since $\mathfrak{M}$ is normal, A and its conjugate[9] $B = (t, 2^{s-1}, b_{12}, b_{22})$ both lie in the generating coset $A\mathfrak{N}'$. Moreover since $\mathfrak{S}_B \nsubseteq \mathfrak{N}_{s-1}$, we have (theorem 6.7) $b_{12} = 2^{s-1}\ (2^s)$. Hence (note 1, lemma 6.8) the generating coset $A\mathfrak{N}'$ contains also $(t, 2^{s-1}, 2^{s-1}, (3^j + 2^{2(s-1)})/t)$. There is one necessary restriction on t, namely (theorem 6.7) $t - (3^j + 2^{2(s-1)})/t = 2^{s-1}\ (2^s)$, i.e. $t^2 - 3^j = 2^{s-1}\ (2^s)$. Hence $t = \pm 3^{j/2}(1 + 2^{s-2})\ (2^r)$, and the generating coset $A\mathfrak{N}'$ contains D or $-D$ or both, where

$$D = (3^{j/2}(1 + 2^{s-2}), 2^{s-1}, 2^{s-1}, (3^j + 2^{2(s-1)})/(1 + 2^{s-2})3^{j/2}).$$

SUFFICIENCY. We can prove that all the groups we have set up are actually normal by showing that the commutators of M, N, D with W_{12}, W_{21}, $V_1^{(3)}$, $V_1^{(-1)}$ are in $\mathfrak{N}'$.

THEOREM 11.6. *Let the first associated subgroup $\mathfrak{N}'$ be given, $s = s(\mathfrak{N}') > 0$. Necessary and sufficient conditions that there exist a normal subgroup $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}'$, $s(\mathfrak{M}) = s$, $\mathfrak{M}/\mathfrak{N}'$ cyclic of preassigned order $u = 2^\gamma > 1$, are: if $s > 2$: $\gamma < r - 2$; if $s = 2$, $r > 2$: $\gamma < r - 2$ if $(\mathfrak{N}')'' = \mathfrak{J}_2$; $\gamma \leqq r - 2$ otherwise; if $s = r = 2$: impossible; if $s = 1$, $r > 2$: $\gamma \leqq r - 2$; if $s = 1$, $r = 2$: $u = 2$. In case $s = 1$, $r > 2$, $(\mathfrak{N}')'' = \mathfrak{N}$, $\mathfrak{N}_1$, and in this case only, there is just one group $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}'$, $\mathfrak{M}/\mathfrak{N}'$ of type $(\gamma, 1)$ for each $1 \leqq \gamma \leqq r - 2$.*

REMARK: The number of groups $\mathfrak{M}$, when there are any, depends on the character of $(\mathfrak{N}')''$ and the presence or absence of $-\mathbf{1}$ in $\mathfrak{N}'$, as explained in the proof.

PROOF: If $s > 2$, then (theorem 11.4) $\mathfrak{M} = \{\mathfrak{N}', A\}$, where $s(A) \geqq s$, $|A| = 3^j\ (2^r)$, $1 < j = 2^{r-2-\gamma} \leqq 2^{r-2}$, $A^u \in \mathfrak{N}'$, $u = 2^\gamma$. Mod $\mathfrak{J}_s$, A must have one of the forms diag $(\pm 3^{j/2}, \pm 3^{j/2})$, diag $(\pm 3^{j/2}(1 + 2^s), \pm 3^{j/2}/(1 + 2^s))$. Hence if $u > 1$ and $(\mathfrak{N}')'' = \mathfrak{J}_s$, there are two or four groups $\mathfrak{M}$ with $o\{\mathfrak{M}/\mathfrak{N}'\} = u$, $\mathfrak{M}' = \mathfrak{N}'$, according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$. If $u > 1$ and $(\mathfrak{N}')'' = \mathfrak{N}_s$, there are one or two such groups according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$.

Mod $\mathfrak{H}_{s+1}$, A must have one of the forms diag $(\pm 3^{j/2}, \pm 3^{j/2})$, $(\pm 3^{j/2}, 2^s, 0, \pm 3^{j/2})$, $(\pm 3^{j/2}, 0, 2^s, \pm 3^{j/2})$, $(\pm 3^{j/2}, 2^s, 2^s, \pm(3^j + 2^{2s})/3^{j/2})$. If $(\mathfrak{N}')'' = \mathfrak{H}_{s+1}$, $s(\mathfrak{N}') = s$, then $\mathfrak{N}' = \{\mathfrak{H}_{s+1}, M\}$ or $\{\mathfrak{H}_{s+1}, -M\}$ or $\{\mathfrak{H}_{s+1}, M, -\mathbf{1}\}$ (theorem 11.3). Since $\mathfrak{S}_A \supseteq \mathfrak{J}_s$ in the last six cases, there are $(u > 1)$ one or two groups $\mathfrak{M}$ with $o\{\mathfrak{M}/\mathfrak{N}'\} = u$, $\mathfrak{M}' = \mathfrak{N}'$ according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$.

Now let $s = 2$, $r > 2$. Then (theorem 11.4) $\mathfrak{M} = \{\mathfrak{N}', A\}$ where $s(A) \geqq 2$, $\Delta = \det A = (-3)^j\ (2^r)$, $1 \leqq j = 2^{r-2-\gamma} \leqq 2^{r-2}$, $A^u \in \mathfrak{N}'$, $u = 2^\gamma$. Mod $\mathfrak{J}_2$, A must have one of the forms diag $(\pm 1, \pm\Delta)$, diag $(\pm 5, \pm\Delta/5)$. $W_{21}AW_{21}^{-1}A^{-1}$ is in $\mathfrak{N}_2$ but not in $\mathfrak{J}_2$ unless $\Delta = 1$ (8). Therefore when $(\mathfrak{N}')'' = \mathfrak{J}_2$, u must be a divisor of 2^{r-3}, and if in addition $u > 1$, there are two or four groups $\mathfrak{M}$

[9] The matrix B is obtained from the p. n. form of A by transformation with W_{12}^m.

with $\mathfrak{M}' = \mathfrak{N}'$, $o\{\mathfrak{M}/\mathfrak{N}'\} = u$ according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$. When $(\mathfrak{N}')'' = \mathfrak{N}_2$, there are one or two groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}'$, $o\{\mathfrak{M}/\mathfrak{N}'\} = u$ according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$.

Mod $\mathfrak{H}_3$, A must have one of the forms diag $(\pm 1, \pm\Delta)$, $(\pm 1, 4, 0, \pm\Delta)$, $(\pm 1, 0, 4, \pm\Delta)$, $(\pm 1, 4, 4, \pm(\Delta + 16))$. But $\mathfrak{S}_A \supseteq \mathfrak{J}_2$ except when A is diag $(\pm 1, \pm\Delta)$ $(\Delta = 1\ (8))$ or $(\pm 1, 4, 4, \pm 13)$. Hence if $\mathfrak{N}' = \mathfrak{H}_3$, there are one or two groups $\mathfrak{M}$ with $o\{\mathfrak{M}/\mathfrak{N}'\} = u$ (u a divisor >1 of 2^{r-2}), $\mathfrak{M}' = \mathfrak{N}'$ according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$.

If $s = r = 2$, then $\mathfrak{M} = \mathfrak{N}'$ follows from $s(\mathfrak{M}) = s(\mathfrak{N}')$, $\mathfrak{M}' = \mathfrak{N}'$.

Now let $s = 1$. Then (theorem 11.4) $\mathfrak{M} = \{\mathfrak{N}', A\}$ or $\{\mathfrak{N}', A, B\}$, where $s(A) > 0$, $\det A = \pm 3^j\ (2^r)$ $(1 \leqq j = 2^{r-2-\gamma} \leqq 2^{r-2})$ and $s(B) > 0$, $\det B = -1\ (2^r)$. Mod $\mathfrak{J}_1$, $\pm A$ must have the form diag $(1, \Delta)$. $W_{21}AW_{21}^{-1}A^{-1}$ is in $\mathfrak{N}_1$, but not in $\mathfrak{J}_1$ unless $\Delta = 1\ (4)$. Therefore if $(\mathfrak{N}')'' = \mathfrak{J}_1$, $\det A = (-3)^j$ $(1 \leqq j = 2^{r-2-\gamma} \leqq 2^{r-2})$. If $r \geqq 3$ and $(\mathfrak{N}')'' = \mathfrak{J}_1$ there are one or two groups $\mathfrak{M} = \{\mathfrak{N}', A\}$ with $o\{\mathfrak{M}/\mathfrak{N}'\} = 2^\gamma$ $(1 \leqq \gamma < r - 2)$, $\mathfrak{M}' = \mathfrak{N}'$, according as $-\mathbf{1} \in \mathfrak{N}'$ or $-\mathbf{1} \notin \mathfrak{N}'$. $\{\mathfrak{N}', B\}$ always contains $\mathfrak{H}_2$ or $\mathfrak{N}_1$. If $r = 2$, there is one group $\{\mathfrak{N}', A\}$ when $\det A = 3$ or $1\ (2^r)$.

If $(\mathfrak{N}')'' = \mathfrak{N}_1$ then $-\mathbf{1} \in \mathfrak{N}'$, and the groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}'$ are in 1, 1 correspondence with the subgroups $\mathfrak{T}$ of the group of odd numbers (2^r), $\mathfrak{M}/\mathfrak{N}' \sim \mathfrak{T}$.

LEMMA 11.7. *If $r > 1$ the cosets of $\mathfrak{G}/\mathfrak{K}$ of determinant Δ are represented by* $\pm G_\Delta = \text{diag}\ (\pm 1, \pm\Delta)$; $\pm H_\Delta = (0, \mp\Delta, \pm 1, 0)$.

PROOF: No two of the above four are in the same coset. The order of $\mathfrak{G}/\mathfrak{K}$ is the quotient $4 \cdot 2^{r-1}$ of the orders of these groups.

THEOREM 11.8. *If $\mathfrak{M}' = \mathfrak{K}$, then $\mathfrak{M}/\mathfrak{K}$ is cyclic, and its order u divides 2^{r-2}. For each such preassigned divisor $u > 1$ of 2^{r-2} there are two groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{K}$, $o\{\mathfrak{M}/\mathfrak{K}\} = u$.*

PROOF: Since $V_1^{(-1)}H_\Delta V_1^{(-1)}H_\Delta^{-1} = -H_\Delta H_\Delta^{-1} = -\mathbf{1}$ is not in $\mathfrak{K}$, hence $\mathfrak{S}_{\pm H} \supseteq \mathfrak{F}$. Moreover, $W_{21}G_\Delta W_{21}^{-1}G_\Delta^{-1}$ is in $\mathfrak{N}_1$ but not in $\mathfrak{K}$ unless $\Delta = 1\ (4)$. Hence the determinants (2^r) in $\mathfrak{M}$ are $(-3)^{\epsilon i}$, $\epsilon = 2^{r-2}/u$, $i = 1, 2, \cdots, u$. $\mathfrak{M} = \{\mathfrak{K}, G_\Delta\}$ or $\{\mathfrak{K}, -G_\Delta\}$ are both normal, and if $\Delta = (-3)^\epsilon \neq 1\ (2^r)$ they both satisfy $\mathfrak{M}' = \mathfrak{K}$.

THEOREM 11.9. *For each preassigned divisor $u = 2^\gamma > 1$ of 2^{r-2} there are four normal groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{F}$ and $\mathfrak{M}/\mathfrak{F}$ of type $(\gamma, 1)$; and four normal groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{F}$ and $\mathfrak{M}/\mathfrak{F}$ of type (γ). When $r = 2$, $\mathfrak{F}$, $\{\mathfrak{F}, \text{diag}\ (1, -1)\}$, and $\{\mathfrak{F}, (0, 1, 1, 0)\}$ are the only groups having $\mathfrak{F}$ for a first associated.*

PROOF: The cosets of $\mathfrak{G}/\mathfrak{F}$ of determinant Δ are represented by G, H of lemma 11.7. The determinants in $\mathfrak{M} = \{\mathfrak{F}, G_\Delta \text{ or } H_\Delta, G_{-1} \text{ or } H_{-1}\}$ $(\Delta = 3^\epsilon)$ are $\pm 3^{\epsilon i}$, $\epsilon = 2^{r-2}/u$, $i = 1, 2, \cdots, u$, so that $\mathfrak{M}/\mathfrak{F}$ is of type $(\gamma, 1)$ and further $\mathfrak{M}' = \mathfrak{F}$. The determinants in $\mathfrak{M} = \{\mathfrak{F}, G_{\pm\Delta} \text{ or } H_{\pm\Delta}\}$ $(\Delta = 3^\epsilon)$ are $3^{\epsilon i}$ (or $(-3^\epsilon)^i$), ϵ, i as above, so that $\mathfrak{M}/\mathfrak{F}$ is of type (γ) and further $\mathfrak{M}' = \mathfrak{F}$. All these groups are normal.

THEOREM 11.10. *The normal groups $\mathfrak{M}$ with $\mathfrak{M}' = \mathfrak{N}_0$ are in 1, 1 correspondence with the subgroups of the multiplicative group of odd numbers (2^r).*

12. **The normal subgroups when $n > 2$, $p = 2$.**

PRELIMINARY CONSIDERATIONS. We have proved that when $n > 2$, if $\mathfrak{N}$ is any normal subgroup of $\mathfrak{G}$, and if $s = s(\mathfrak{N})$, then $\mathfrak{N} \supseteq \mathfrak{N}_s$. Obviously also, $\mathfrak{N}_s^{*\,5a} \supseteq \mathfrak{N}$, so we describe all normal subgroups when we describe all normal subgroups $\mathfrak{N}$ satisfying $\mathfrak{N}_s^* \supseteq \mathfrak{N} \supseteq \mathfrak{N}_s$, $s = s(\mathfrak{N})$. A convenient description is again afforded through the groups $\mathfrak{N}'$ consisting of all the matrices of $\mathfrak{N}$ with determinant 1 (2^r). We shall describe all possible (normal in $\mathfrak{G}$) groups $\mathfrak{N}'$, and later all possible normal subgroups $\mathfrak{N}$ of $\mathfrak{G}$. The description will analyze $\mathfrak{N}/\mathfrak{N}'$ and $\mathfrak{N}'/\mathfrak{N}_s$; also a convenient algorithm will appear for telling what matrices $\mathfrak{N}$ contains.

LEMMA 12.1. $3^w = 1 + 2^s\ (2^{s+2})$ $(s \geqq 3,\ w = 2^{s-2})$.

LEMMA 12.2. *All the odd numbers* mod 2^s $(s \geqq 3)$ *are the first* 2^{s-2} *powers of* 3, *and their negatives.*

ALGORITHM 12.3. *If n is odd, $\mathfrak{N}' = \mathfrak{N}_s$ $(s = s(\mathfrak{N}))$. If $n = 2^\lambda n' > 1$ (n' odd, $\lambda > 0$), then if $s \leqq 1$, $\mathfrak{N}' = \mathfrak{N}_s$; if $s \geqq 2$, there exists a matrix A in $\mathfrak{N}'$ with the following properties*: (a) *A is of order* $2^{\mu-k}$ mod $\mathfrak{N}_s$, $0 \leqq \mu - k \leqq \mu = \min(\lambda, s-2)$; (b) *The answer to "Is B in $\mathfrak{N}'$?" is the same as the answer to "Is $A^z B$ [or $\pm A^z B$] in $\mathfrak{N}_s$ for any value of z: $1 \leqq z \leqq 2^{\mu-k}$?"*

We merely remark that (a) and (b) of algorithm 12.3 are necessary and sufficient conditions that $\mathfrak{N}' = \{\mathfrak{N}_s, A\}$ [or $\{\mathfrak{N}_s, A, -\mathbf{1}\}$], where $s(A) \geqq s$, $\det A = 1\ (2^r)$, and A is of order $2^{\mu-k}$ mod $\mathfrak{N}_s$. Hence we need not prove algorithm 12.3 if we prove the equivalent

THEOREM 12.4. *If n is odd, $\mathfrak{N}' = \mathfrak{N}_s$. If $n = 2^\lambda n'$ (n' odd, $\lambda > 0$), then if $s \leqq 1$, $\mathfrak{N}' = \mathfrak{N}_s$; if $s \geqq 2$, the groups $\mathfrak{N}'$ (for given s) are in 1, 1 correspondence with the subgroups of the group of odd numbers* mod $2^{\min(\lambda+2,s)}$ [*i.e. the Abelian group $\mathfrak{Q}_1$ of order $2^{\min(\lambda+1,s-1)}$ and type* $(1, \min(\lambda, s-2))$]. *The correspondence is such that given $s \geqq 2$ then to each subgroup $\mathfrak{T}_1$ of $\mathfrak{Q}_1$, there corresponds a group $\mathfrak{N}'(s(\mathfrak{N}') = s$; all matrices of $\mathfrak{N}'$ have determinant* 1 $(2^r))$ *such that $\mathfrak{N}'/\mathfrak{N}_s \sim \mathfrak{T}_1$.*

PROOF: (a) If n is odd, let A be any matrix of $\mathfrak{N}'$ and let t be its multiplier (2^s). Since $\det A = 1\ (2^r)$, we have $t^n = 1\ (2^s)$; therefore $t = 1\ (2^s)$. (b) If n is even and $s \leqq 1$, $\mathfrak{N}' \subseteq \mathfrak{N}_s$ from the definitions of these two groups. Hence $\mathfrak{N}' = \mathfrak{N}_s$. If n is even and $s \geqq 2$, consider $\mathfrak{N}_s^{*\prime}$ consisting of all matrices A with $s(A) \geqq s$; $\det A = 1\ (2^r)$. Any multiplier $t\ (2^s)$ of a matrix of $\mathfrak{N}_s^{*\prime}$ satisfies $t^n = 1\ (2^s)$, $t^\zeta = 1\ (2^s)$ $(\zeta = 2^\lambda)$; further any odd number $t\ (2^s)$ satisfying $t^\zeta = 1\ (2^s)$ is the multiplier of some matrix A of $\mathfrak{N}_s^{*\prime}$, viz.: diag $(t + l2^s, t, t, \cdots, t)$, where l satisfies $t^n + t^{n-1} l 2^s = 1\ (2^r)$. These numbers $t\ (2^s)$ form the group $\mathfrak{Q}_1$: for generators we may take $t = 3^k$ $(k = 2^{s-2-\mu})$, $t = -1$. Thus $\mathfrak{N}_s^{*\prime}/\mathfrak{N}_s \sim \mathfrak{Q}_1$.

Now $\mathfrak{N}'$ is a subgroup of $\mathfrak{N}_s^{*\prime}$; the multipliers of the matrices of $\mathfrak{N}'$ moreover form a subgroup of the multipliers of the matrices of $\mathfrak{N}_s^{*\prime}$: the multiplier of the product of two matrices is the product of their multipliers. From this it is clear that the multipliers of $\mathfrak{N}'$ form a subgroup $\mathfrak{T}_1$ of $\mathfrak{Q}_1$; hence the correspondence $\mathfrak{N}'/\mathfrak{N}_s \sim \mathfrak{T}_1$. Thus the groups $\mathfrak{N}'$ have been completely described and classified, and we have given an algorithm for telling what matrices $\mathfrak{N}'$ contains.

We turn now to a discussion of the groups $\mathfrak{N}$. If $r = 1$, $\mathfrak{N} = \mathfrak{N}' = \mathfrak{N}_s$. Let $r > 1$.

THEOREM 12.5. *If n is odd, or if $s \leqq 1$, then for given s the groups $\mathfrak{N}(s(\mathfrak{N}) = s)$ are in 1, 1 correspondence with the subgroups of the Abelian group $\mathfrak{Q}$ of order 2^{r-1} and type $(1, r-2)$; the correspondence is such that to each subgroup $\mathfrak{T}$ of $\mathfrak{Q}$ there corresponds a group $\mathfrak{N}$ $(s(\mathfrak{N}) = s)$ such that $\mathfrak{N}/\mathfrak{N}_s \sim \mathfrak{T}$. An algorithm similar to algorithm 12.3 may be stated.*

PROOF: $\mathfrak{Q}$ is the group of odd numbers $\pm 3^k$ (2^r). Now $\mathfrak{N}_s^*/\mathfrak{N}_s \sim \mathfrak{Q}$ in case n is odd. For let h satisfy $hn = k$ (2^{r-2}); then $\mathfrak{N}_s^*$ contains in fact a matrix $\pm\mathbf{3}^h$ of determinant $\pm 3^k$ (2^r). Matrices of the same (2^r) determinant belong in the same coset, since their quotient has determinant 1 (2^r).

In case $s \leqq 1$, the matrix diag $(1 + 2l, 1, 1, \cdots, 1)$, where l satisfies $2l = -1 \pm 3^k$ (2^r), has determinant $\pm 3^k$ (2^r), and is in $\mathfrak{N}_s^*$.

THEOREM 12.6. *Let n be even: $n = 2^\lambda n'$ (n' odd, $\lambda > 0$). Then if $s(\mathfrak{N}) = s \geqq 2$, $\mathfrak{N}/\mathfrak{N}'$ is cyclic.*

PROOF: Since n is even, the determinants of $\mathfrak{N}$ are all quadratic residues mod 2^s. Hence if $s \geqq 3$ there is no determinant $= -3^\epsilon$ (2^r) for any positive ϵ. If $s = 2$, there is no determinant $3^\epsilon(2^r)$ (ϵ odd). There might be a matrix of determinant -3^ϵ (ϵ odd) in $\mathfrak{N}$ however. In this case $\mathfrak{N}$ would contain matrices of all determinants 3^ϵ (for any even ϵ) (see (a) of the proof of lemma 12.7), so that $\mathfrak{N}/\mathfrak{N}'$ is generated by the coset of determinant -3 (2^r). (There is in this case a coset of determinant -3.)

LEMMA 12.7. *There is a matrix A in a generating coset of $\mathfrak{N}/\mathfrak{N}'$ $(s(A) \geqq s = s(\mathfrak{N}) \geqq 2)$ whose determinant has the following character*: (a) *if $s \geqq 3$,* $\det A = 3^\theta$, $2^\mu \leqq \theta = 2^\gamma \leqq 2^{r-2}$, $\mu = \min(\lambda, s-2)$; (b) *if $s = 2$,* $\det A = (-3)^\theta$, $1 \leqq \theta = 2^\gamma \leqq 2^{r-2}$.

PROOF: (a) Let B be in a generating coset, and let its determinant be $3^{\epsilon\theta}$, ϵ odd, $\theta = 2^\gamma$. Then a suitable power $A = B^\sigma$ of B has determinant 3^θ and generates B (mod $\mathfrak{N}'$). Suppose $A = (a_{ij})$. We have $3^\theta = \det A = a_{11}^n$ (2^s), so that 3^θ is the 2^λ power of an odd number (2^s). Hence $\gamma \geqq \mu$. Statement (b) says essentially that no matrix of $\mathfrak{N}$ has determinant -3^ϵ (ϵ even) or 3^ϵ (ϵ odd).

Lemma 12.7 contains necessary restrictions on the determinant of a matrix A, which with $\mathfrak{N}'$ generates $\mathfrak{N}$. When we have found necessary conditions on the multiplier of A, we shall have a set of necessary and sufficient number-theoretic conditions for the existence of a group $\mathfrak{M}$ containing a given group $\mathfrak{N}'$ $(\mathfrak{M}' = \mathfrak{N}')$ with preassigned index.

Since $\mathfrak{N} = \mathfrak{N}'$ when $\gamma = r - 2$, we shall assume $\mu \leqq \gamma < r - 2$ in case $s \geqq 3$; $0 < \gamma < r - 2$ in case $s = 2$. There are three possibilities for the multipliers (2^s) of $\mathfrak{N}'$ when $s \geqq 3$:

i) $\{3^{iu}\}$, $i = 1, 2, \cdots, 2^{\mu-k}$, $u = 2^{s-2-\mu+k}$, a cyclic group of order $2^{\mu-k}$ $(0 \leqq k \leqq \mu)$.

ii) $\{\pm 3^{iu}\}$, the above and their negatives.

iii) $\{(-3^u)^i\}$, $i = 1, 2, \cdots, 2^{\mu-k}$, a cyclic group of order $2^{\mu-k}$ $(0 \leqq k < \mu)$.

When $s = 2$, the multipliers of $\mathfrak{N}'$ are i) 1 or ii) ± 1.

THEOREM 12.8. ($s \geqq 3$.) *Let* $\gamma < s - 2$. ($s = s(\mathfrak{N}') \geqq 3$.) *Necessary and sufficient conditions that there exist a group* $\mathfrak{M}$ *such that* $\mathfrak{M}' = \mathfrak{N}'$, $\mathfrak{M}/\mathfrak{N}'$ *is of order* $2^{r-2-\gamma}$ ($\mu \leqq \gamma < r - 2$) *are:* $r \geqq s + k$ *if* $\mathfrak{N}'$ *is of type* i) *or* ii); $r > s + k$ *if* $\mathfrak{N}'$ *is of type* iii). *The number of distinct groups* $\mathfrak{M}$ *for fixed* $\mathfrak{N}'$ *and fixed* $\mathfrak{M}/\mathfrak{N}'$, *if there are any, is:* 2^{k+1} *if* $\mathfrak{N}'$ *is of type* i) *or* iii); 2^k *is* $\mathfrak{N}'$ *if of type* ii).

PROOF: NECESSITY. Let $\mathfrak{M}$ exist. Then A exists such that $s(A) \geqq s(\mathfrak{N}')$; A is of order $z = 2^{r-2-\gamma}$ mod $\mathfrak{N}'$. Let the multiplier (2^s) of A be $t = \pm 3^{\delta v}$ (δ odd, $0 < \delta v = \delta 2^{\delta_1} \leqq 2^{s-2}$). Since A^z is in $\mathfrak{N}'$, we see that t^z must be one of the multipliers of $\mathfrak{N}'$. From this and $\delta_1 \geqq 0$, the conditions of the theorem follow.

SUFFICIENCY. We have to prove that if our conditions hold, then for any $\gamma < s - 2$ and $\mu \leqq \gamma < r - 2$ ($s = s(\mathfrak{N}')$), there exists a matrix A with $s(A) \geqq s$, A of order $z = 2^{r-2-\gamma}$ mod $\mathfrak{N}'$.

If A exists and is of order $2^{r-2-\gamma}$ mod $\mathfrak{N}'$, then $\det A = 3^{\epsilon\theta}$, ϵ odd, $\theta = 2^\gamma$. Hence a suitable power $B = A^\sigma$ ($\sigma\epsilon = 1\ (2^{r-2-\gamma})$) of A has $\det B = 3^\theta\ (2^r)$, and B generates A mod $\mathfrak{N}'$; B is of order $2^{r-2-\gamma}$ mod $\mathfrak{N}'$. Thus the groups $\mathfrak{N}$ such that $\mathfrak{N}/\mathfrak{N}'$ is cyclic of order $2^{r-2-\gamma}$ ($\mu \leqq \gamma < r - 2$, $s \geqq 3$) are in 1, 1 correspondence with the cosets $A\mathfrak{N}'$, where A is a variable matrix with

$$s(A) \geqq s, \det A = 3^\theta\ (2^r), A^z \epsilon \mathfrak{N}'\ (z = 2^{r-2-\gamma}). \tag{1}$$

Let A have multiplier $t = \pm 3^{\delta v}\ (2^s)$ (δ odd, $0 < \delta v = \delta 2^{\delta_1} \leqq 2^{s-2}$), and let t satisfy $\det A = 3^\theta\ (2^s)$, i.e.

$$\delta_1 = \gamma - \lambda\ (\geqq 0), \qquad \delta n' = 1\ (2^{s-2-\gamma}). \tag{2}$$

Then $t^n = 3^\theta\ (2^s)$, so that the matrix $A = \text{diag}\ (t + l2^s, t, t, \cdots, t)$ satisfies (1) if l satisfies $t^n + lt^{n-1}2^s = 3^\theta\ (2^r)$. In fact $s(A) \geqq s$, $\det A = 3^\theta$, and the multiplier (2^s) of the $2^{r-2-\gamma}$ power of A is 3^ν, $\nu = \delta 2^{\delta_1 + r - 2 - \gamma}$. Since $\mu = \lambda$ if ($\mu \leqq$) $\gamma < s - 2$, theorem 12.8 gives a sufficient condition that this be one of the numbers i) ii), iii).

The *number* of eligible multipliers $t\ (2^s)$ (satisfying (2)) for matrices A is easily seen to be $2^{\lambda+1}$; they are $t = \pm 3^{\delta v}$, $v = 2^{\delta_1}$, $\delta_1 = \gamma - \lambda$, $\delta = \bar{\delta} + i2^{s-2-\gamma}$ ($i = 1, 2, \cdots, 2^\lambda$), where $\bar{\delta} n' = 1\ (2^{s-2-\gamma})$. Each eligible[10] multiplier actually leads moreover to a diagonal matrix of determinant $3^\theta\ (2^r)$, with $s(A) \geqq s$. The coset to which A belongs contains matrices which are of the same determinant (2^r) and same order mod $\mathfrak{N}'$ as A, and whose multipliers (2^s) form a collection of $2^{\lambda-k}$, $2^{\lambda-k+1}$, $2^{\lambda-k}$ distinct (2^s) numbers, according as $\mathfrak{N}'$ is of type i), ii), or iii). Thus the $2^{\lambda+1}$ diagonal matrices A we have constructed satisfying (1) fall into sets (depending on the group $\mathfrak{N}'$) such that $\{\mathfrak{N}', A\}$ is the same for each of the $2^{\lambda-k}$ or $2^{\lambda-k+1}$ matrices A in a given set. The number of distinct groups $\mathfrak{N}$ is therefore the quotient of $2^{\lambda+1}$ by either $2^{\lambda-k}$ or $2^{\lambda-k+1}$, Q.E.D.

THEOREM 12.9. ($s \geqq 3$.) *Let* $s - 2 \leqq \gamma < r - 2$ ($s = s(\mathfrak{N}') \geqq 3$). *There*

[10] What we are doing here in effect is considering cosets $A\mathfrak{N}_s$.

always exist one or more groups $\mathfrak{M}$ *such that* $\mathfrak{M}' = \mathfrak{N}'$, $\mathfrak{M}/\mathfrak{N}'$ *is of order* $2^{r-2-\gamma}$. *If* $\kappa = \min(\mu, \mu - k + r - 2 - \gamma)$, *when* $\mathfrak{N}'$ *is of type* i) *or* ii); $\kappa = \min(\mu, \mu - k + r - 2 - \gamma - 1)$, *when* $\mathfrak{N}'$ *is of type* iii), *then the number of distinct groups* $\mathfrak{M}$ *is:* $2^{\kappa+\mu-k+1}$ *if* $\mathfrak{N}'$ *is of type* i) *or* iii); $2^{\kappa+\mu-k}$ *if* $\mathfrak{N}'$ *is of type* ii).

PROOF: As in theorem 12.8, we seek a matrix A such that $\det A = 3^\theta$, $s(A) \geqq s$, A is of order $2^{r-2-\gamma}$ mod $\mathfrak{N}'$. Let the multiplier of A be $t = \pm 3^{\delta v}$ (2^s) (δ odd, $0 < \delta v = \delta 2^{\delta_1} \leqq 2^{s-2}$). Then t is (2^s) a power of 3^π ($\pi = 2^{s-2-\kappa}$). For we must have $t^n = 3^\theta = 1$ (2^s), $\theta = 2^\gamma$, i.e. $\delta_1 \geqq s - 2 - \lambda$. Furthermore t^z, $z = 2^{r-2-\gamma}$, must be a multiplier (2^s) of a matrix of $\mathfrak{N}'$, i.e.

$$\delta_1 + (r - 2 - \gamma) \geqq s - 2 - (\mu - k)$$

if $\mathfrak{N}'$ is of type i) or ii);

$$\delta_1 + (r - 2 - \gamma) \geqq s - 2 - (\mu - k) + 1,$$

if $\mathfrak{N}'$ is of type iii). Hence $\delta_1 \geqq s - 2 - \kappa$, as stated. Furthermore, if t is any one of the numbers $\pm 3^{\pi i}$ ($\pi = 2^{s-2-\kappa}$, $i = 1, 2, \cdots, 2^\kappa$), there is a matrix A with multiplier t (2^s) such that $\det A = 3^\theta$, $s(A) \geqq s$, A is of order $2^{r-2-\gamma}$ mod $\mathfrak{N}'$. (See (3) and the above inequalities.)

Thus, since as in theorem 12.8, the $2^{\kappa+1}$ eligible matrices A we have constructed fall into sets (depending on the group $\mathfrak{N}'$) such that $\{\mathfrak{N}', A\}$ is the same for each of the $2^{\mu-k}$ or $2^{\mu-k+1}$ matrices in a given set, therefore the number of distinct groups $\mathfrak{N}$ is the quotient of $2^{\kappa+1}$ by either $2^{\mu-k}$ or $2^{\mu-k+1}$, Q.E.D.

We have now only the case n even, $s = 2$, $r > 2$ to consider. As we have seen, the multipliers (mod 4) of $\mathfrak{N}'$ are in this case 1 or ± 1. $\mathfrak{N}/\mathfrak{N}'$ is cyclic, and generated by a coset which contains a matrix $A(s(A) \geqq s)$ such that $\det A = -3$, or $= 3^\theta$ (2^r) $(2 \leqq \theta = 2^\gamma < 2^{r-2})$.

THEOREM 12.10. ($s = 2$.) *Let* $s(\mathfrak{N}') = 2$. *There always exist groups* $\mathfrak{M}$ ($\mathfrak{M}' = \mathfrak{N}'$) *containing* $\mathfrak{N}'$ *with index* $2^{r-2-\gamma}$ $(0 \leqq \gamma < r - 2)$; *namely: two such groups if* $\mathfrak{N}'$ *is of type* i); *one such group if* $\mathfrak{N}'$ *is of type* ii).

PROOF: We seek A: $s(A) \geqq s$, $\det A = (-3)^\theta$ (2^r), $\theta = 2^\gamma$, A of order $2^{r-2-\gamma}$ mod $\mathfrak{N}'$, such that $\mathfrak{M} = \{\mathfrak{N}', A\}$. If t is ± 1 (4), then

$$A = \operatorname{diag}(t + 4l, t, t, \cdots, t),$$

$t^n + 4lt^{n-1} = (-3)^\theta$ (2^r), is soluble for l; from each of the two odd numbers (4) we get a matrix A, from which fact theorem 12.10 follows.

13. **Relations among A^p: miscellaneous results** ($p > 2$). In this section we prove a simple theorem about the structure of $\mathfrak{G}$.

LEMMA 13.1. *If* a_{jj} *is prime to* p, *a prime* > 2, *and* $a_{ii} = a_{jj}$ (p^u) $(u > 0)$, $a_{ii} \not\equiv a_{jj}$ (p^{u+1}), *then* $a_{ii}^p = a_{jj}^p$ (p^{u+1}), $a_{ii}^p \not\equiv a_{jj}^p$ (p^{u+2}); *hence*

$$([a_{ii}^p - a_{jj}^p]/[a_{ii} - a_{jj}], p^2) = p.$$

THEOREM 13.2. *The* pth *power of an* s*-matrix is an* $(s + 1)$*-matrix* $(0 < s < r)$ $(p > 2)$.

PROOF: FIRST STEP: $s(A^p) \geqq 1 + s(A)$ if $0 < s(A) < r$. We form the pth power $B = (b_{ij})$ of $A = (a_{ij})$ mod p^{2s}, and find

$$b_{ij} = a_{ij}[a_{ii}^{p-1} + a_{ii}^{p-2}a_{jj} + \cdots + a_{jj}^{p-1}] \ (p^{2s}) \qquad (i \neq j);$$

$b_{ij} = a_{ii}^p \ (p^{2s})$ $(i = j)$. In view of the preceding lemma, the first step is obvious, because that lemma says that the long parenthesis is divisible by p.

SECOND STEP: $s(A^p) \leqq 1 + s(A)$ if $0 < s(A) < r$. Since $s(B^{-1}AB) = s(A)$ and since $(B^{-1}AB)^p = B^{-1}A^pB$ for any matrix B in $\mathfrak{G}$, we may assume without loss of generality that A is in p.n. form; then the second step becomes obvious.

COROLLARY: *The order of an s-matrix is* λp^{r-s}, $(0 < s \leqq r)$, $\lambda \mid \phi(p^{\max([(r+1)/2],s)})$.

THEOREM 13.3. *The center of* $\mathfrak{G}$ *consists of all its r-matrices (scalar matrices).*

PROOF: If A is in the center, then it is identical with all its conjugates. Hence it is in pseudo-normal form. Then $p^s = 0 \ (p^r)$; otherwise transformation by W_{12} would actually change A in the (1, 1)th place. Hence A is an r-matrix. On the other hand, every r-matrix is in the center.

THEOREM 13.4. *The group* $\mathfrak{D}_s$ *generated by the commutators of the* p^s *powers of matrices of* $\mathfrak{G}$ *contains* $\mathfrak{N}_s$.

PROOF: Pick α so that $\alpha \not\equiv 0$, $\alpha^w \not\equiv 1 \ (p^r)$ $(w = p^s)$. Then

$$[W_{ij}^w V_j^{(v)} W_{ij}^{-w} V_j^{(1/v)}]^{1/(1-v)} = W_{ij}^w \qquad (v = \alpha^w).$$

Previous investigations of the group $\mathfrak{G}$ have been fragmentary, and confined to the case $r = 1$. A bibliography of them appears in the Duke University Journal, 1935, p. 27, footnote. Material slightly relevant to the present discussion is also found in M. M. Feldstein's paper in Trans. Amer. Math. Soc. 25 (1923), p. 224.

UNIVERSITY OF MINNESOTA.

ANNALS OF MATHEMATICS
Vol. 39, No. 2, April, 1938

QUADRATIC NULL FORMS OVER A FUNCTION FIELD

BY A. ADRIAN ALBERT

(Received November 17, 1937)

1. Introduction

A quadratic form

$$f = f(x_1, \cdots, x_m) = \sum_{i,j=1}^{m} a_{ij} x_i x_j, \tag{1}$$

with coefficients a_{ij} in a field K, is called a null form over K if there exist quantities $x_{10}, \cdots, x_{m0}$ in K and not all zero such that $f(x_{10}, \cdots, x_{m0}) = 0$. One of the most interesting results on rational quadratic forms is the Meyer theorem[1] stating that every indefinite f in five or more variables is a null form. This result was generalized by H. Hasse[2] to forms over an algebraic number field.

There are infinite modular fields with arithmetic properties very much like those of the rational number field but actually somewhat simpler. These are the simple transcendental extensions F of finite fields L. The fields corresponding to algebraic number fields are now algebraic extensions K of finite degree over F (called algebraic function fields of one indeterminate and constant field L). It is then true that every quadratic form over K in five or more variables is a null form. This result was given by E. Witt[3] in a paper on quadratic forms over a field of characteristic not two, and the reader was referred to papers by three other writers[4] for proof. It is clear that the proof is also valid for L of characteristic two.[5] However it seems to depend essentially on the finiteness of L.

---

[1] For a proof using concepts analogous to those used here see H. Hasse, Journal für Mathematik, vol. 152 (1923), pp. 129–48.

[2] See the Journal für Math., vol. 153 (1924), pp. 113–30.

[3] Journal für Math., vol. 176 (1937), pp. 31–44.

[4] C. Chevalley, Abh. Math., Sem. Hamburg, vol. 11 (1935), pp. 73–5; E. Warning, Abh. same volume, pp. 76–83; C. Tsen, *Algebren über Funktionenkorpern*, Dissertation, Gottingen, 1934.

[5] The reasoning by which the null form theorem is derivable from the aforesaid references is non-trivial and it is regrettable that it does not appear in Witt's paper. A reconstruction of it was communicated to the author by the referee and follows in a modified form. Let K have degree n and a basis $u_1, \cdots, u_n$ over $F = L(\eta)$, and $f = f(x_1, \cdots, x_5)$ be a form in five variables over K. Tsen's thesis contains a result of Artin which suggests the replacement of the coefficients and variables of f by linear combinations of the u_k with coefficients in F. Then $f = 0$ is equivalent to $f_k = 0$ $(k = 1, \cdots, n)$, where the f_k are n quadratic forms over F in $5n$ variables y_j. There is clearly no loss of generality if we assume that the coefficients of f_k are polynomials in η of maximum degree s. We then write $y_j = z_{j0} + z_{j1}\eta + \cdots + z_{jt}\eta^t$ with $5n(t+1)$ variables z_{ji} over L. All the f_k are now polynomials in η of degree at most $2t + s$, and we set the coefficients of the powers of η equal to zero. They are

We shall sketch a proof here for the case where L is finite of characteristic not two which is very much like the proof of Hasse[2] for forms over an algebraic number field. It depends on the arithmetic theorems of E. Witt[6] on normal division algebras and may be useful as a method for other types of fields. It is not valid without revision for fields of characteristic two and it is precisely this point that lends interest to this type of proof. We also give an analysis of the quadratic null form theory over arbitrary fields of characteristic two.

The reduction theory of quadratic forms over a field of characteristic two is essentially different from that for all other characteristics, and our null form theorem actually has a far simpler proof in this case. We make the result depend upon the elementary algebraic theorem which states that the adjunction of the square roots of all of its quantities to K may be accomplished by the adjunction of a single square root. This result as well as our theorem holds for L any perfect field of characteristic two. In fact let L be any perfect field of characteristic p, $F = L(\eta_1, \cdots, \eta_r)$ be the field of all rational functions of r independent indeterminates η_j over L, K be algebraic of finite degree over F, $q = p^e$. Then the field G of the q-th roots of all quantities of K is shown to have degree $\rho = p^{er}$ over K. We use this to show that every form

$$a_1 x_1^q + \cdots + a_m x_m^q \qquad (a_i \text{ in } K) \tag{2}$$

is a null form over K if $m > \rho$. However there exist forms (2) which are not null forms for every $m \leqq \rho$. The result is applied for $r = e = 1$, $p = 2$, and gives our quadratic null form theorem by a trivial argument.

2. Arithmetic theory

A simple transcendental extension F of an arbitrary field L is the set $F = L(\eta)$ of all rational functions with coefficients in L of an indeterminate η over L. The polynomials in η form a ring $J = L[\eta]$ whose quotient field is F. This ring is a principal ideal ring whose units are the non-zero elements of L, whose primes are the polynomials irreducible in L.

If π is a prime of J and α is in F we may write

$$\alpha = \pi^r a b^{-1} \tag{3}$$

where a and b are in J and prime to π. If the rational integral *order* ν of α is non-negative we say that α is integral for π. If m is in J we say that α is integral for m if it is integral for all prime factors of m.

---

at most $2t + s + 1$ quadratic forms over L in $5n(t+1)$ variables, and by Chevalley's theorem applied for polynomials all of degree two have a simultaneous non-trivial solution if $5n(t+1) > 2n(2t+s+1)$, that is $t > 2s - 3$. Hence f is a null form. A similar argument shows that if we replace F by $L(\eta_1, \cdots, \eta_r)$ for independent indeterminates $\eta_1, \cdots, \eta_r$ over a finite field L, and if f is any polynomial of degree q in $x_1, \cdots, x_m$ such that $f(0, \cdots, 0) = 0$, then $f = 0$ has another solution in K if $m > q^{r+1}$.

[6] Math. Annalen, vol. 110 (1934), pp. 12–28.

By a *congruence*

$$\alpha \equiv \beta \pmod{m} \qquad (\alpha, \beta \text{ in } F, m \text{ in } J) \tag{4}$$

we mean that α and β are integral for m and that the order of $\alpha - \beta$ is at least that of m for every prime factor of m. Our congruences may then be easily shown to have all the usual properties of elementary number-theoretic congruences.

The defining quantity η of F over L is not unique. In particular $F = L(\lambda)$, $\lambda = 1/\eta$. The general element of F has the form $\alpha = ab^{-1}$ with a and b in $L[\eta]$. If a has degree s in η and b has degree t it is easily seen that the order of α with respect to the prime λ of $J_\lambda = L[\lambda]$ is the difference $t - s$. We then consider congruences $\alpha \equiv \beta \pmod{\lambda^k m}$ where m is in J and α and β are integral for λ as well as for m. Then it is not difficult to prove the Chinese Remainder Theorem.

LEMMA 1. *Let $m_1, \cdots, m_t$ be quantities of J which are relatively prime in pairs, $m_0 = \lambda^k$ for a positive integer k, $\alpha_0, \cdots, \alpha_t$ be in F such that α_i is integral for m_i $(i = 0, \cdots, t)$. Then there exists an α in F which is integral for all the m_i and is such that*

$$\alpha \equiv \alpha_i \pmod{m_i} \qquad (i = 0, \cdots, t). \tag{5}$$

We assume the result above and now also that L is a finite field. Let K be algebraic of finite degree over $F = L(\eta)$. The integers of K have been described and form a ring H_η in which the principal results of the algebraic number ideal theory hold.[7] Here again we have another ring H_λ, $\lambda = 1/\eta$ and have prime ideal divisors of λ in H_λ. Lemma 1 may be applied to prove a final Chinese Remainder

THEOREM 1. *Let $m_1, \cdots, m_t$ be ideals of H_η which are relatively prime in pairs, α_i in K be integral for m_i $(i = 1, \cdots, t)$. Assume that $m_{t+1}, \cdots, m_s$ are ideals of H_λ relatively prime in pairs and such that every prime ideal divisor in H_λ of $\mu = m_{t+1} \cdots m_s$ divides λ, α_j in K is integral for m_j $(j = t + 1, \cdots, s)$. Then there exists a γ in K such that*

$$\gamma \equiv \alpha_i \pmod{m_i} \qquad (i = 1, \cdots, s). \tag{6}$$

The Chinese Remainder Theorem is customarily used to combine congruences and it is rather surprising, in view of the large number of papers on the arithmetic of fields K of our type,[8] that it has not been proved before. It may be applied to prove

LEMMA 2. *Let $P_1, \cdots, P_t$ be a finite number of ideals which are either prime*

[7] Cf. Chapter XIV of Van der Waerden's Moderne Algebra, vol II, Berlin, 1931.

[8] References may be found in Witt's paper of footnote 6, as well as in the article giving the culminating existence theorems of the theory by T. Tannaka, Science reports of the Tohoku Imperial University, vol. 24 (1935), pp. 165–72. Tannaka is of course able to combine his results with those of Witt to show that all normal division algebras, over an algebraic function field of one indeterminate and finite constant field, are cyclic.

ideals of H_η or prime ideal divisors of $\lambda = 1/\eta$ and α, β be in K such that $\alpha\beta \neq 0$. Then there exist positive rational integers k_i and quantities x and y in K such that

$$\gamma = \alpha x^2 + \beta y^2 \tag{7}$$

is integral for all the P_i and for all $i = 1, \cdots, t$ the quantity γ is not congruent to a square modulo $P_i^{k_i}$.

The valuation theory of our field K is well known. If ϕ is any valuation of K there is a unique derived field K_ϕ whose elements are power series in an integral quantity π divisible by a prime ideal P of the corresponding integral[9] domain H of K but not by P^2. Lemma 2 then has as immediate consequence

LEMMA 3. *Let $\phi_1, \cdots, \phi_t$ be a finite number of inequivalent valuations of K, K_{ϕ_i} the corresponding derived fields, α and β be in K such that $\alpha\beta \neq 0$. Then there exists a quantity $\gamma = \alpha x^2 + \beta y^2$ (x, y in K) such that $K_{\phi_i}(\sqrt{\gamma})$ has degree two over K_{ϕ_i} for $i = 1, \cdots, t$.*

3. Quadratic null forms

The theory of quadratic null forms over a field K of characteristic not two has a very interesting connection with quaternion algebras.[10] Suppose that f is a quadratic form in n variables over K so that there is no loss of generality if we write

$$f = a_1 x_1^2 + \cdots + a_n x_n^2. \tag{8}$$

If some $a_i = 0$ we put $x_i = 1$, $x_j = 0$ for $j \neq i$ and have $f = 0$, f is a null form. Hence let all the a_i be non-zero.

Clearly $f = a_1 x_1^2$ is never a null form. Also $a_1 x_1^2 + a_2 x_2^2$ is a null form if and only if $-a_1 a_2$ is the square of a quantity of K. We now temporarily restrict our attention to

$$f = a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 \qquad (a_i \neq 0 \text{ in } K). \tag{9}$$

This f is a null form if and only if

$$g = df = a x_1^2 + b x_2^2 - ab x_3^2 \tag{10}$$

is a null form where

$$d = -(a_1 a_2)^{-1} a_3, \qquad a = -a_2^{-1} a_3, \qquad b = -a_1^{-1} a_3. \tag{11}$$

This is the critical point of the theory and, as the author observed,[10] we have

LEMMA 4. *The form (9) is a null form if and only if the corresponding generalized quaternion algebra*

$$Q = (1, i, j, ij), \qquad i^2 = a, j^2 = b, ji = -ij \tag{12}$$

is a total matric algebra.

---

[9] The domain H is the ring of all quantities of K which are roots of equations with leading coefficient unity and other coefficients polynomials in the generating indeterminate over L. Clearly H varies with η. The indeterminate η is always chosen so that K is separable over $F = L(\eta)$, and then possibly replaced by $\lambda = 1/\eta$.

[10] Bulletin of the American Mathematical Society, vol. 39 (1933), pp. 585–8. See also Witt's paper in footnote 3 for a reference to this connection as given by Artin in lectures at Hamburg. The precise results are stated in Witt's Theorems 13 and 14.

We now consider the form (10). As has been stated in many places, a quadratic field $K(\sqrt{\gamma})$ is equivalent to a subfield of the algebra Q of (12) if and only if $\gamma = ax_1^2 + bx_2^2 - abx_3^2$ for x_1, x_2, x_3 in K. When Q is a total matric algebra this is always possible. If Q is a division algebra so that (10) is not a null form this is true if and only if $K(\sqrt{\gamma})$ splits Q. We apply this to (8) and write

$$g = df = (ax_1^2 + bx_2^2 - abx_3^2) - (c_4x_4^2 + \cdots + c_nx_n^2). \tag{13}$$

Thus f is a null form if and only if there exists a quantity of the form

$$\gamma = c_4x_4^2 + \cdots + c_nx_n^2, \tag{14}$$

with

$$c_i = (a_1a_2)^{-1}a_3a_i \qquad (i = 4, \cdots, n) \tag{15}$$

such that $K(\sqrt{\gamma})$ splits Q.

We apply this result to the case of an algebraic function field K of one indeterminate and finite constant field L of characteristic $p \neq 2$. Also let Q be a division algebra. By the results of Witt[6] there is a finite number of valuations $\phi_1, \cdots, \phi_t$ such that the scalar extensions $Q_{K_{\phi_i}}$ of Q is a division algebra. By the same results a field $K(\sqrt{\gamma})$ splits Q if and only if $K_{\phi_i}(\sqrt{\gamma})$ has degree two over K_{ϕ_i} for $i = 1, \cdots, t$. Apply Lemma 3 with $\alpha = c_4$, $\beta = c_5$ and obtain x_4, x_5 in K such that $K(\sqrt{\gamma})$ splits Q for $\gamma = c_4x_4^2 + c_5x_5^2$. Then g is a null form and so is $f = a_1x_1^2 + \cdots + a_5x_5^2$. Thus we have given a proof (without details) of the case of characteristic not two of the main

THEOREM 2. *Every quadratic form in five of more variables over an algebraic function field of one indeterminate and finite constant field is a null form.*

There remains the case where the coefficient field has characteristic two. We shall give a complete and simple proof of certain results on pure forms in this case. Theorem 2 for the case of characteristic two will then be a trivial corollary.

4. Adjunction of q-th roots to a field

Our results on forms will be consequences of certain properties of fields of characteristic p. The property which will be given as Theorem 5 is an immediate consequence[11] of results already derived by the author in the case of the field of Theorem 2. However we shall derive the more general results both because of the resulting theorems on forms which will be given here, and because of certain consequent theorems on normal division algebras which will be derived in a later paper.

If $\eta_1, \cdots, \eta_r$ are independent indeterminates over a field L and K is algebraic of finite degree over $F = L(\eta_1, \cdots, \eta_r)$ we call K an algebraic function field of r indeterminates over its constant field L. The indeterminates η_i are

[11] See Theorems 1 and 2 of the author's Bulletin paper, vol. 43 (1937), pp. 733-6.

not unique, but the degree of transcendency r is uniquely determined by K and L. We now prove[12]

THEOREM 3. *Let K be an algebraic function field of r indeterminates over a perfect constant field L of characteristic p. Then K is separable of finite degree over $F = L(\eta_1, \cdots, \eta_r)$ for properly chosen indeterminates η_j.*

The result is known[11] for $r = 1$ and we make an induction on r and assume our result true for $r - 1$. Let K_0 be the maximum separable subfield over $F = L(\eta_1, \cdots, \eta_r)$ of K and let $G = K_0(y)$, $y^p = a$ in K_0. We are clearly assuming that $K_0 < K$, $a \neq b^p$ for any b of K_0. The field $F(y)$ has degree mp over F, where m is the degree of $F(a)$ over F. Also, y is a root of its minimum function of degree mp and coefficients in F. We multiply this minimum function by a suitably chosen quantity of F and obtain a polynomial

$$(16) \quad f = f(\eta, \eta_1, \cdots, \eta_r) = a_0\eta^{mp} + a_1\eta^{(m-1)p} + \cdots + a_{m-1}\eta^p + a_m$$

whose coefficients a_i are in the polynomial ring $L[\eta_1, \cdots, \eta_r]$ such that unity is their greatest common divisor. Evidently f is an irreducible polynomial of the ring $L[\eta, \eta_1, \cdots, \eta_r]$, where $\eta, \eta_1, \cdots, \eta_r$ are independent indeterminates over L. In particular this implies that not all of the a_i are in $L[\eta_1^p, \cdots, \eta_r^p]$ and hence we may assume that

$$(17) \quad f = f(\eta_1) = b_0\eta_1^t + b_1\eta_1^{t-1} + \cdots + b_t,$$

with b_j in $L[\eta, \eta_2, \cdots, \eta_r]$, and $b_j \neq 0$ for at least one integer j not divisible by p. If $f(\eta_1)$ were reducible in the field $L(\eta, \eta_2, \cdots, \eta_r)$ of its coefficients then it is known[13] that f would be a reducible polynomial of $L[\eta, \eta_1, \cdots, \eta_r]$, a contradiction.

Assume first that $y, \eta_2, \cdots, \eta_r$ are independent indeterminates over L. Then $f(y, \eta_1, \cdots, \eta_r) = 0$. But by the proof above $f(y, \eta_0, \eta_2, \cdots, \eta_r)$ is irreducible in $L[y, \eta_2, \cdots, \eta_r]$ and is the minimum function of η_1. Hence $F_1(\eta_1)$ is separable of finite degree t over $F_1 = L(y, \eta_2, \cdots, \eta_r)$. Since K_0 is separable over $L(\eta_1, \cdots, \eta_r)$ the composite $K_0(y)$ of K_0 and the field $L(y, \eta_1, \cdots, \eta_r)$ is separable over $L(y, \eta_1, \cdots, \eta_r)$, and hence separable over $L(y, \eta_2, \cdots, \eta_r)$ as desired.

There remains the case where $y, \eta_2, \cdots, \eta_r$ are not independent indeterminates over L. Then y is algebraic over $L(\eta_2, \cdots, \eta_r)$ and by the hypothesis of our induction $L(y, \eta_2, \cdots, \eta_r)$ is separable of finite degree over $L(\xi_2, \cdots, \xi_r)$. Clearly $L(y, \eta_1, \cdots, \eta_r)$ is then separable of finite degree over $L(\eta_1, \xi_2, \cdots, \xi_r)$ and, as before, $K_0(y)$ is separable over $L(\eta_1, \xi_2, \cdots, \xi_r)$ as desired.

The field K is obtained from K_0 in a finite number of steps. At each stage

[12] Theorems 4 and 5 may be proved without the use of Theorem 3, but the proof of Theorem 5 is somewhat more complicated. This was done by the author originally and the consequent results on forms obtained. The result of Theorem 3 for $r = 1$ has been known for several years, and the possibility of its proof for arbitrary r was suggested to the author by Dr. S. MacLane. The proof given here is the result of our collaboration.

[13] Cf. Lemma 3 of Section 2.9 in the author's *Modern Higher Algebra*, Chicago, 1937.

we have a field K_i which is separable of finite degree over $L(\eta_{1i}, \cdots, \eta_{ri})$ and have $K_{i+1} = K_i(y_{i+1})$, $y_{i+1}^p = a_{i+1}$ in K_i. By our proof K_{i+1} is separable of finite degree over $L(\eta_{1,i+1}, \cdots, \eta_{r,i+1})$. Hence the field K is separable of finite degree over $L(\eta_{1s}, \cdots, \eta_{rs})$ as desired.

We use the

DEFINITION. *An algebraic extension G of a field K of characteristic p is called a Kummer field of exponent e over K if $q = p^e$, g^q is in K for every g of G, at least one y of G has the property y^t is not in K for $t = 1, \cdots, q-1$.*

It is trivial to show that our definition is equivalent to the statement that every quantity g of G defines a simple Kummer extension $K(g)$ of degree $p^f \leqq p^e$ over K, and that at least one such extension has degree p^e over K. Observe that we are not assuming that G has finite degree over K. But we shall prove

THEOREM 4. *Let K be separable of finite degree over $L(\eta_1, \cdots, \eta_r)$ with the η_i independent indeterminates over the perfect field L of characteristic p. Then the field*

$$K(y_1, \cdots, y_r), \qquad y_i^q = \eta_i, \ q = p^e, \tag{18}$$

has degree p^{er} and exponent q over K and contains (in the sense of equivalence) every Kummer extension of exponent q over K.

For let $F = L(\eta_1, \cdots, \eta_r)$ so that $K = F(\theta)$. Here θ is a root of a separable irreducible polynomial of degree n with coefficients in F. If $F(\theta) > F(\theta^q)$ then $F(\theta)$ would be inseparable over $F(\theta^q)$ and hence over F. This is contrary to hypothesis. Hence $F(\theta) = F(\theta^q)$ for every power q of p. It follows that $1, \theta^q, \cdots, \theta^{q(n-1)}$ are linearly independent in F and that every quantity of K has the form

$$a = a_0 + a_1\theta^q + \cdots + a_{n-1}(\theta^q)^{n-1} \qquad (a_i \text{ in } F). \tag{19}$$

But L is perfect of characteristic p and

$$\sum_{i_j=1}^{n_j} a_{i_1\cdots i_r}\eta_1^{i_1}\cdots\eta_r^{i_r} = \left(\sum_{i_j=1}^{n_i} b_{i_1\cdots i_r}y_1^{i_1}\cdots y_r^{i_r}\right)^q \tag{20}$$

for all $a_{i_1\cdots i_r}$ in L and $b_{i_1\cdots i_r}$ in L such that

$$a_{i_1\cdots i_r} = (b_{i_1\cdots i_r})^q. \tag{21}$$

It follows that every a of K has the form $a = b^q$ with b in $K(y_1, \cdots, y_r)$. This implies that $K(y_1, \cdots, y_r)$ contains all Kummer extensions of exponent q over K. If

$$\sum_{j_k=0}^{q-1} c_{j_1\cdots j_r}y_1^{j_1}\cdots y_r^{j_r} = 0 \tag{22}$$

for $c_{j_1\cdots j_r}$ in K and not all zero, then

$$\sum_{j_k=0}^{q-1} c_{j_1\cdots j_r}^q\eta_1^{j_1}\cdots\eta_r^{j_r} = 0. \tag{23}$$

Write $c_{j_1 \cdots j_r}$ as a linear combination of $1, \theta, \cdots, \theta^{n-1}$ with coefficients in F and have its q-th power a linear combination of $1, \theta^q, \cdots, \theta^{q(n-1)}$ with coefficients in $L[\eta_1^q, \cdots, \eta_r^q]$. The linear independence of the powers of θ^q above in F implies that

$$\sum_{j_k=0}^{q-1} d_{j_1 \cdots j_r} \eta_1^{j_1} \cdots \eta_r^{j_r} = 0 \tag{24}$$

with $d_{j_1 \cdots j_r}$ in $L[\eta_1^q, \cdots, \eta_r^q]$ and not all zero. This is not possible, since the quantities $\eta_1, \cdots, \eta_r$ are independent indeterminates over F.

We may combine Theorems 3 and 4 and have immediately

THEOREM 5. *Let K be an algebraic function field of r indeterminates and perfect constant field L. Then there exists an algebraic extension of degree p^{er} and exponent $q = p^e$ which contains the q-th root of every quantity of K.*

5. Application to pure forms

Consider *pure* forms of degree $q = p^e$ over K of characteristic p, that is, forms

$$f = b_1 x_1^q + \cdots + b_m x_m^q \qquad (b_i \text{ in } K). \tag{25}$$

We call f a null form over K if $f(x_{10}, \cdots, x_{m0}) = 0$ for x_{i0} not all zero and in K, and prove

THEOREM 6. *Let K be an algebraic function field of r indeterminates and perfect constant field L. Then every pure form in $m > q^r$ variables x_i is a null form. However there exist non-null forms for every $m \leqq q^r$.*

For as we have seen in Theorem 5 there exists a field G of degree $q^r = \rho$ over K such that if $u_1, \cdots, u_\rho$ is a basis of G over K then every a of K has the unique form

$$a = (a_1 u_1 + \cdots + a_\rho u_\rho)^q \qquad (a_i \text{ in } K). \tag{26}$$

In particular we may write

$$b_j = \left(\sum_{i=1}^{\rho} b_{ji} u_i \right)^q \qquad (j = 1, \cdots, m) \tag{27}$$

for the coefficients b_j of f and have

$$f = \left(\sum_{j=1}^{m} \sum_{i=1}^{\rho} x_j b_{ji} u_i \right)^q \tag{28}$$

in our field K of characteristic p. Then $f = 0$ for x_i in K if and only if $\sum_{i=1}^{\rho} \left(\sum_{j=1}^{m} x_j b_{ji} \right) u_i = 0$. The linear independence of the u_i in K implies that f is a null form if and only if the homogeneous system

$$\sum_{j=1}^{m} x_j b_{ji} = 0 \qquad (i = 1, \cdots, \rho) \tag{29}$$

has a non-trivial solution. Such a solution exists if $m > \rho$. But if $m \leqq \rho$ we may choose the b_{ji} and hence the coefficients of f so that the matrix (b_{ji})

has maximum rank m, and the only solution of $f = 0$ is the trivial one with all the $x_i = 0$.

The reader should observe that our proof actually gives necessary and sufficient conditions that the form f of (25), (27) be a null form.

6. Quadratic null forms

The case $r = 1$, $q = p = 2$ of Theorem 6 may be stated as

THEOREM 7. *Every quadratic form $a_1x_1^2 + a_2x_2^2 + a_3x_3^2$ over an algebraic function field K of one indeterminate and perfect constant field of characteristic two is a null form.*

We now apply this result to forms in $m \geqq 5$ variables. The author has proved[14] that every quadratic form $\sum_{i,j=1}^{m} a_{ij}x_ix_j$ over a field K of characteristic two is equivalent in K to

$$(30) \qquad f = a_1x_1^2 + \cdots + a_mx_m^2 + x_1x_2 + x_3x_4 + \cdots + x_{2r-1}x_{2r}.$$

Here $2r$ is the rank of the skew matrix $A - A'$, $A = (a_{ij})$. We put $m = 5$ and have the possible values $r = 0, 1, 2$. Thus all forms in five variables are equivalent to one and only one of

$$(31) \quad f_0 = a_1x_1^2 + \cdots + a_5x_5^2, \qquad f_0 + x_1x_2, \qquad f_0 + x_1x_2 + x_3x_4.$$

In every case $f(x_1, 0, x_3, 0, x_5) = ax_1^2 + a_3x_3^2 + a_5x_5^2$ is a null form by Theorem 7 and hence so is f. We have proved

THEOREM 8. *Let K be an algebraic function field of one indeterminate over a perfect field of characteristic two. Then every quadratic form in five or more variables over K is a null form.*

The theory of quadratic null forms over K in $m \leqq 4$ variables may again be connected with quaternion algebras. Theorem 7 first implies also

THEOREM 9. *All quaternary forms*

$$(32) \quad f = a_1x_1^2 + \cdots + a_4x_4^2, \qquad f = a_1x_1^2 + \cdots + a_4x_4^2 + x_1x_2 \quad (a_i \text{ in } K)$$

over an algebraic function field K of one indeterminate and perfect constant field of characteristic two are null forms.

To complete the theory we temporarily assume that K is an *arbitrary* field of characteristic two and make some general remarks. If in (30) any $a_i = 0$ we put the corresponding $x_i = 1$ and all the remaining $x_j = 0$ and have $f = 0$, f is a null form. Hence let all the $a_j \neq 0$. If $m = 1$ the form $f = a_1x_1^2$ is never a null form.

The form $f = a_1x_1^2 + \cdots + a_mx_m^2$ is a null form if and only if the quantities

---

[14] In a paper on symmetric matrices over a field of characteristic 2 which will appear in the Transactions of the American Mathematical Society, May, 1938.

$\sqrt{a_1}, \cdots, \sqrt{a_m}$ are linearly dependent in K. For $f = \left(\sum_{i=1}^{m} a_i^{\frac{1}{2}} x_i\right)^2 = 0$ for x_i in K not all zero if and only if $\sum_{i=1}^{m} a_i^{\frac{1}{2}} x_i = 0$. Thus we need only consider forms (30) with $r \neq 0$.

If $f = a_1 x_1^2 + x_1 x_2 + a_2 x_2^2$ then f if a null form if and only if the equation $y^2 + y + a_1 a_2 = 0$ is reducible. This is a condition of the same type as that for forms in two variables over a field of characteristic not two and we cannot analyze this case further.

We now let Q be a normal simple algebra of degree two over K. Then it is known[15] that

$$(33)\quad Q = (1, i, j, ij), \qquad i^2 = i + a, \qquad j^2 = b, \qquad ji = (i+1)j,$$

with a and b in K. The general element of Q is $x = \xi_1 + \xi_2 i + \xi_3 j + \xi_4 ij$ and its minimum function is

$$(34)\qquad \phi(\omega) = \omega^2 - \xi_2 \omega + N(x) = 0,$$

where we use $-1 = 1$ in K and write[15]

$$(35)\qquad N(x) = \xi_1^2 + \xi_1 \xi_2 + \xi_2^2 a + b(\xi_3^2 + \xi_3 \xi_4 + \xi_4^2 a).$$

If Q is a division algebra every quantity x of Q with ξ_2, ξ_3, ξ_4 not all zero generates a subfield of Q. This field is inseparable if and only if $\xi_2 = 0$.

If Q is not a division algebra then it is a total matric algebra and contains an $x \neq 0$ such that $x^2 = 0$. This x has $\xi_2 = 0$ and we have the

LEMMA. *The algebra Q is a division algebra if and only if the form*

$$(36)\qquad \xi_1^2 + b(\xi_3^2 + \xi_3 \xi_4 + \xi_4^2 a)$$

is not a null form.

In case the equation $i^2 = i + a$ is irreducible in K the algebra $K[i]$ is a cyclic field and the criterion of our lemma is simply that b shall not be the norm $\xi_3^2 + \xi_3 \xi_4 + \xi_4^2 a$ of any $\xi_3 + \xi_4 i$ in $K[i]$. Observe also that the form (36), which we shall call the first reduced norm form of Q, is the square of the general x of Q whose square is in K. We then treat the only remaining case of quadratic forms with $m = 3$ in

THEOREM 10. *A ternary form $f = a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 + x_1 x_2$ is a null form if and only if the corresponding algebra* (33) *with*

$$(37)\qquad a = a_1 a_2, \qquad b = (a_1 a_3)^{-1}$$

is a total matric algebra.

---

[15] Cf. the author's paper in the American Journal of Mathematics, vol. 56 (1934), p. 78.

For clearly f is a null form if and only if

$$g = a_3^{-1}f = x_3^2 + (a_1a_3)^{-1}[(a_1x_1)^2 + (a_1x_1)x_2 + (a_1a_2)x_2^2] \tag{38}$$

is a null form.

We next consider forms with $m = 4$ and $r > 0$. If $r = 1$ then we may write $f = a_1x_1^2 + \cdots + a_4x_4^2 + x_1x_2$ and have f a null form if and only if $a_3^{-1}f$ is a null form, that is

$$(a_3^{-1}a_4)x_4^2 = x_3^2 + (a_1a_3)^{-1}[(a_1x_1)^2 + (a_1x_1)x_2 + (a_1a_2)x_2^2]. \tag{39}$$

Applying our discussion of Q we have

THEOREM 11. *Let $f = a_1x_1^2 + \cdots + a_4x_4^2 + x_1x_2$ with $a_i \neq 0$ in K define an algebra Q by* (37), (33). *Then f is a null form if and only if either Q is a total matric algebra, or $a_3^{-1}a_4$ is the square of a quantity in K, or the field $K(\sqrt{a_3^{-1}a_4})$ is equivalent to an inseparable subfield of the division algebra Q.*

We next let $f = a_1x_1^2 + \cdots + a_4x_4^2 + x_1x_2 + x_3x_4$. Form the expression

$$a_3^{-1}f = y_3^2 + b(y_1^2 + y_1y_2 + ay_2^2) + y_3y_4 + cy_4^2, \tag{40}$$

with $y_3 = x_3$, $y_1 = a_1x_1$, $y_2 = x_2$, $y_4 = a_3^{-1}x_4$, and

$$a = a_1a_2, \qquad b = (a_1a_3)^{-1}, \qquad c = a_3^{-3}a_4. \tag{41}$$

If the corresponding algebra (33) is a total matric algebra we make $f = 0$ by putting $y_4 = 0$ and making $y_3^2 + b(y_1^2 + y_1y_2 + ay_2^2) = 0$. Otherwise $f = 0$ implies that $y_4 \neq 0$ and f is a null form if and only if

$$\xi_1^2 + \xi_1 + a + b(\xi_3^2 + \xi_3\xi_4 + \xi_4^2a) = c + a. \tag{42}$$

The left member of (42) is the second reduced norm form of Q. It is the norm of the general quantity of Q satisfying the so called normed equation $\omega^2 = \omega + N(x)$, and is obtained from (35) by putting $\xi_2 = 1$. We have proved

THEOREM 12. *The form $f = a_1x_1^2 + \cdots + a_4x_4^2 + x_1x_2 + x_3x_4$ is a null form if and only if either the algebra Q of* (33), (41) *is a total matric algebra, or the field $K(x)$, $x^2 = x + c + a$, is equivalent to a (separable) subfield over K of Q.*

Quadratic forms in $m > 4$ variables over an arbitrary field of characteristic two may be written as $f = f_4 + f_{40}$ where f_4 is a form in four variables, f_{40} is a form in the remaining $m - 4$ variables. Moreover either f has $r = 1$ or we may assume that f_4 has $r = 2$. In the former case f is a null form if and only if there exists a quantity of the form $\rho = a_3^{-1}(a_4x_4^2 + \cdots + a_mx_m^2)$ such that $K(\sqrt{\rho}) = K$ or is equivalent to an inseparable subfield of Q. In the latter case f is a null form if and only if $c + a + f_{40}$ represents a quantity d such that $K(x)$, $x^2 = x + d$, is equivalent to a separable subfield over K of Q. We now return[16] to the special field K of Theorem 9.

---

[16] The results of Witt on quadratic forms of a field of characteristic not two may probably be obtained for the characteristic two case only for forms with cross product terms. It would be very interesting to study the analogues of Witt's results for our characteristic two case but the author has not yet done so.

Observe that the second form of (32) in Theorem 9 is a form considered in Theorem 11 and is a null form because there is only a single inseparable quadratic field over a field K of the type considered in Theorem 9. Note also that Theorem 12 implies that, by the results of Witt,[6] the form f is a null form if and only if the degrees of $K_{\phi_i}(x)$ defined by $x^2 = x + c + a$ and over K_{ϕ_i} are two for every valuation ϕ_i of K such that the algebra Q, with coefficient field extended to be K_{ϕ_i}, is a division algebra over K_{ϕ_i}.

THE UNIVERSITY OF CHICAGO.